

**ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA**

DIPARTIMENTO DI FISICA E ASTRONOMIA

**Dottorato di Ricerca in Fisica
Ciclo XXVI**

**MEASUREMENTS OF $t\bar{t}$ DIFFERENTIAL
CROSS SECTION AT THE ATLAS
EXPERIMENT IN pp COLLISIONS AT
 $\sqrt{s} = 7\text{ TeV}$ AND $\sqrt{s} = 8\text{ TeV}$**

Tesi di PhD di: Matteo Franchini

Coordinatore di Dottorato:

Prof. Fabio Ortolani

Relatore:

Chiar.mo Prof. Antonio Zoccoli

Correlatori:

Dott. Matteo Negrini

Dott. Roberto Spighi

Settore Concorsuale di afferenza: 02/A1

Settore Scientifico disciplinare: FIS/04

Esame Finale – Anno 2014

Contents

Introduction.....	1
1 The top quark.....	7
1.1 The Standard Model.....	7
1.1.1 Quantum Electrodynamics.....	8
1.1.2 Quantum Chromodynamic.....	10
1.1.3 The Weak Interaction.....	11
1.1.4 Electroweak unification.....	14
1.1.5 Spontaneous symmetry breaking.....	16
1.2 Top Quark.....	19
1.2.1 Top quark decay	20
1.2.2 Top Quark pair production.....	21
1.2.3 $t\bar{t}$ pair cross section studies	27
1.2.4 Differential $t\bar{t}$ cross section	31
1.2.5 Single top quark production.....	38
1.2.6 Top quark mass.....	39

2 The Large Hadron Collider and the ATLAS Detector.....	43
2.1 LHC	43
2.2 Introduction to the ATLAS detector	46
2.3 Magnetic system	47
2.4 The Inner Detector.....	49
2.4.1 Pixel detector.....	50
2.4.2 Semi Conductor Tracker.....	51
2.4.3 Transition Radiation Tracker	52
2.5 Calorimetric System	52
2.5.1 Electromagnetic calorimeter	53
2.5.2 Hadronic calorimeters.....	55
2.6 Muon spectrometer	56
2.7 Luminosity Measurement.....	59
2.7.1 Lucid.....	59
2.7.2 BCM	60
2.7 Minimum Bias Trigger Scintillator.....	60
2.8 Trigger System.....	61
 3 Data samples and Monte Carlo models for Signal and	
Background.....	65
3.1. Detector Data Samples	65
3.2. Monte Carlo Simulation.....	69
3.3. $t\bar{t}$ Monte Carlo Signal.....	71
3.4. Monte Carlo and Data Driven Background sources.....	72
3.4.1 QCD or Fake Leptons background.....	76
3.4.2 W+jets background.....	79

4 Object Reconstruction	83
4.1 Jets.....	83
4.2 Electrons	88
4.3 Muons.....	91
4.4 Missing Transverse Energy.....	94
4.5 Boosted Objects.....	96
4.6 Jets substructure observables.....	99
4.7 Jet grooming algorithms.....	101
4.8 HEPTopTagger	105
4.9 Template Overlap Method.....	108
4.10 TOM efficiency and rejection power.....	110
 5 Analysis Method and Event Selection	 121
5.1 Event Selection	123
5.1.1 Resolved Analysis	123
5.1.2 Boosted Analysis	124
5.2 Physical objects definition	126
5.2.1 Resolved Analysis	126
5.2.2 Boosted Analysis	128
5.3 Data/MC comparison plots	131
5.4 Unfolding.....	137
5.4.1 Unfolding methods.....	138
5.4.2 Unfolding Studies.....	143
5.4.3 Parton and Particle Level.....	146
5.4.4 Migration Matrices.....	148
5.4.5 Unfolding Error Propagation.....	152
5.4.6 Channel combination	153
5.5 Analysis framework	153

6 Results	157
6.1 Statistic uncertainties	158
6.2 Systematic uncertainties	158
6.2.1 Signal modeling	159
6.2.2 Background modeling	162
6.2.3 Detector and reconstruction modeling	164
6.3 Results of the 2011 analysis	166
6.3.1 Differential cross section as a function of the top quark p_T : $\frac{1}{\sigma} \frac{d\sigma}{dp_T^t}$	168
6.3.2 Differential cross section as a function of $t\bar{t}$ mass: $\frac{1}{\sigma} \frac{d\sigma}{dm_{t\bar{t}}}$	171
6.3.3 Differential cross section as a function of $t\bar{t}$ transverse momentum: $\frac{1}{\sigma} \frac{d\sigma}{dp_T^{t\bar{t}}}$	173
6.3.4 Differential cross section as a function of $t\bar{t}$ rapidity: $\frac{1}{\sigma} \frac{d\sigma}{dy^{t\bar{t}}}$	175
6.4 Results of the 2012 analysis	178
 Conclusion	 185
 Appendix A	 187
Appendix B	195

Introduction

The top quark has been discovered in 1995 by the CDF and D0 experiments at the Tevatron proton-antiproton ($p\bar{p}$) collider. Since its discovery, the study of the top quark has represented one of the most interesting and investigated fields in particle physics because of its very peculiar properties, like the high mass and the short life-time. The top quark mass value $m_t = 173.2 \pm 0.9 \text{ GeV}$ [7] constitutes one of the Standard Model (SM) fundamental free parameters and makes the top quark the heaviest known fundamental particle. For this reason, in many beyond SM theories, the top quark is the preferred coupling partner for most of the predicted new particles, as the Z' boson.

The top quark decays via the $t \rightarrow Wq$ process, where the produced quark q has bottom flavor in almost every cases; this implies that the $|V_{tb}|$ element of the Cabibbo-

Kobayashi–Maskawa (CKM) matrix is close to one. This establishes an important experimental result for the SM. The top quark has also the peculiarity to decay before its hadronization. This implies the unique possibility to observe the properties of a *bare quark*, as for example the spin effects on the decaying products.

Top quark studies play an important role in the physics program of the Large Hadron Collider (LHC) and in particular for the ATLAS experiment. Thanks to the very high luminosity and collision energy at LHC, the number of produced top quarks is considerably larger with respect to Tevatron, allowing to perform high-statistic precision measurements.

In proton–proton (pp) collisions, top quarks are produced in pairs or individually through the strong or the weak interaction respectively, allowing important tests on the features of these two fundamental forces included in the SM. The large amount of top quark pairs ($t\bar{t}$) produced, allows measurements of the differential cross section as a function of different kinematic variables (rapidity, momentum, mass, etc.). These studies are used in order to verify the SM predictions and to validate Monte Carlo models. They also have a great importance in the characterization of the top quark production mechanism. It is a key ingredient for the investigation of physical channels where the top quark is a dominant background (as for in some Higgs boson analysis) and permits to search for new physics by looking at the presence of new resonances decaying into $t\bar{t}$ pairs. Moreover, the measurements of $t\bar{t}$ production, having a multifarious signature involving jets, electrons, muons and missing transverse energy, needs a detailed

understanding of the whole apparatus; it constitutes then a good opportunity to test the comprehension of the detector.

Two $t\bar{t}$ differential cross section measurements, performed applying different event selections on two different data samples, are presented in this thesis.

The first one is the measurement of the $t\bar{t}$ system differential production cross section performed on all the statistics collected in 2011 at a center of mass energy $\sqrt{s} = 7\text{TeV}$, with an integrated luminosity of $\mathcal{L} = 4.6\text{fb}^{-1}$. The analysis has been performed in the *lepton+jets* channel, where one of the W bosons coming from the top quark decays in a lepton and neutrino and the other one into two quarks. A cut-based selection is performed in order to reduce background, than events are unfolded to the *parton level* in order to allow comparison with theoretical predictions and results from other experiments. The differential cross section spectra as a function of the $t\bar{t}$ and top quark kinematic variables are measured and found to be in good agreement with the expectations; these results have been collected in an ATLAS public conference note [76]. During this measurement I followed all the analysis steps and in particular I has been involved in the implementation of the unfolding techniques and in the estimation of the systematic uncertainties.

Thanks to the increased center of mass energy to $\sqrt{s} = 8\text{TeV}$ in pp collisions acquired during 2012, it is possible to study the cross section behavior at an unprecedented momentum scale. Many analyses, also in the top quark sector, have specifically concentrated efforts on the study of these frontier high-energy regions, in order to search for hints of new physics or confirmation of the SM. Many specific

algorithms have been developed in order to deal with such a high-momentum objects that possess peculiar signatures. In particular, in the hadronic decay of high- p_T top quarks, where the W boson decays into two quarks, the final state products can partially overlap and be reconstructed in a single, energetic and large radius jet, called *fat jet* or *large-R jet*. The identification of these boosted objects rely on tagging algorithms based on the *large-R* jets internal substructure. I'm involved in the application and tuning of some of these *tagging* techniques and in particular on a template based one called template overlap method (TOM). In this prospective, the second analysis described in this PhD thesis aims to study the behavior of $t\bar{t}$ pairs at high transverse momentum, measuring the differential production cross section as a function of the hadronic top p_T using a specific cut-based selection exploiting jet substructure properties. Studies for the application of more sophisticated selection algorithms are ongoing and only preliminary studies on the TOM technique will be presented here. The *lepton+jets* events used, have been extracted from the 2012 data samples at a center of mass energy $\sqrt{s} = 8\text{TeV}$ with an integrated luminosity of $\mathcal{L} = 20\text{fb}^{-1}$. Data have been unfolded at the *particle level* in order to obtain a result more compatible with theoretical predictions and other experiments results. This analysis is one of the first performed at such high momentum region and the results presented are still preliminary. I am the main code developer and analyzer for this measurements and I am currently performing all the analysis steps from the selection to the background estimates and the unfolding process. I also became one of the developers of an analysis package included in the ATLAS official analysis framework called Top Root Core (TRC). The package, named TopD3PDBoosted, is

designed to perform full top quark analyses in both boosted and resolved regimes; it is already used in more than one analysis. The package also includes the TOM algorithm ready to be used in boosted selections. I have been appointed as one of the editors of the internal note regarding this analysis [77].

The structure of the thesis is the following. In *Chapter 1* the theoretical aspects of the top quark and its cross section measurements are considered. The description of the ATLAS detector is done in *Chapter 2*. A detailed description of the collected data samples and the Monte Carlo simulations for both $t\bar{t}$ signal and the different backgrounds considered are reported in *Chapter 3*, together with the data-driven background estimation methods. In *Chapter 4* the reconstruction of physical objects used during the analysis is fully described beside the treatment of high transverse momentum objects topology and some of the main techniques adopted in ATLAS for their reconstruction. The events selection, the comparison between data and Monte Carlo events and the unfolding procedure are described in *Chapter 5* while the evaluation of the measure uncertainties and the presentation of the final differential cross section results are accomplished in *Chapter 6*. In the *Conclusion*, a summary of the results obtained is presented.

Chapter 1

The top quark

1.1 The Standard Model

The Standard Model (SM) is a theory that brilliantly includes all the known subnuclear particles and their interactions into a coherent scheme, with the only exception of the gravitational force. The SM is based on a relatively small number of elementary particles (and respective antiparticles) like: leptons, quarks, and the gauge bosons that are the mediators of the different forces (see Figure 1.1); particles and antiparticles have the same mass, but opposite charges. Leptons and quarks are divided in three generations with a strong hierarchy in mass. Each lepton generation is formed by a neutral and a -1 electric charged* particle, while quark generations are composed

* electric charges are expressed in units of the absolute electron charge value $|e|$.

by a pair of $+2/3$ and a $-1/3$ charged particle each. The interactions between particles are described as gauge quantum fields that interact via gauge bosons; the quarks interact via strong, weak and electromagnetic forces, the leptons only via electromagnetic and weak ones, while neutrinos interact only via the weak force.

mass→	2.4 MeV	1.27 GeV	171.2 GeV	0
charge→	$2/3$	$2/3$	$2/3$	0
spin→	$1/2$	$1/2$	$1/2$	1
name→	u up	c charm	t top	γ photon
Quarks	4.8 MeV	104 MeV	4.2 GeV	0
	$-1/3$	$-1/3$	$-1/3$	0
	$1/2$	$1/2$	$1/2$	1
	d down	s strange	b bottom	g gluon
Leptons	<2.2 eV	<0.17 MeV	<15.5 MeV	91.2 GeV
	0	0	0	0
	$1/2$	$1/2$	$1/2$	1
	ν_e electron neutrino	ν_μ muon neutrino	ν_τ tau neutrino	Z^0 Z boson
	0.511 MeV	105.7 MeV	1.777 GeV	80.4 GeV
	-1	-1	-1	± 1
	$1/2$	$1/2$	$1/2$	1
	e electron	μ muon	τ tau	W^\pm W boson
				Gauge Bosons

Figure 1.1 – Standard model constituent summary.

1.1.1 Quantum Electrodynamics

Quantum Electrodynamics or QED is the simplest gauge field theory and it describes the electromagnetic force; the force mediator is the photon, a massless gauge boson. The other gauge theories of the Standard Model are, in some extent, an extension of it. The QED Lagrangian is:

$$\mathcal{L} = i\bar{\Psi}\gamma^\mu\mathfrak{D}_\mu\Psi - m\bar{\Psi}\Psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} \quad (1.1)$$

where Ψ is the QED quantum field, m is mass constant,

$$\mathfrak{D}_\mu = \partial_\mu + ieA_\mu(x) \quad (1.2)$$

and

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu \quad (1.3)$$

where e is the elementary charge unit, related to the electromagnetic interaction coupling constant α_e through the formula $e = \sqrt{4\pi\alpha_e}$ and A_μ is the electromagnetic vector potential. This Lagrangian is invariant under the local $U(1)$ rotation

$$\Psi = e^{i\alpha(x)}\Psi \quad (1.4)$$

and under the gauge transformation

$$A_\mu \rightarrow A_\mu - \frac{1}{e}\partial_\mu\alpha(x) \quad (1.5)$$

where $\alpha(x)$ is an arbitrarily gauge field. The addition of a non-null mass term as $\frac{1}{2}m_\gamma A^\mu A_\mu$ in Equation 1.1 would violate the request of the gauge symmetry, violating the experimentally observed massless of the photon. The global $U(1)$ symmetry of QED leads to conservation of the electromagnetic charge[1].

1.1.2 Quantum Chromodynamic

The Quantum Chromodynamics (QCD) describes the strong interactions. It introduces a new quantum number called *color*, from which the name chromodynamic, having three possible states – red, blue and green. Color was first introduced to explain the observation of two unexpected experimental results: the existence of degenerate baryon states like the Δ^{++} and the strange value of the K meson decay rate. The additional color quantum number solves the problem of the apparent violation of the Fermi-Dirac statistic in Δ^{++} , a baryon formed by 3 quarks up (uuu) that, without the further. Using the new quantum number in the cross section calculation it also gets rid of the unexpected results in the K decay rate, found to be 3 times smaller than the theoretical predictions.

In QCD, quarks have never been observed as free states but they can be only found in colorless confined states (baryons or mesons): baryons are composed by three valence-quarks in a color singlet state, while mesons are quark/anti-quark systems forming a color/anti-color state.

The gauge theory of QCD is invariant under transformations of the non-Abelian $SU(3)$ group leading to the conservation of the color quantum number. The QCD Lagrangian is given by:

$$\mathcal{L}_{QCD} = i\bar{\Psi}\gamma^\mu(\partial_\mu - ig_s A_\mu^a T_a)\Psi - m\bar{\Psi}\Psi - \frac{1}{2}tr[F_{\mu\nu}F^{\mu\nu}] \quad (1.6)$$

It appears similar to the QED one with the substitution with the proper coupling constant $e \rightarrow g_s = \sqrt{4\pi\alpha_s}$, but with few important differences: instead of just one,

there are eight gauge fields, corresponding to the eight generators of $SU(3)$, the T^a matrices, where

$$T^a = \frac{1}{2} \lambda^a \quad (1.7)$$

λ^a are the Gell-Mann matrices that do not commute, leading to a more complicated $F_{\mu\nu}$ tensor than in the corresponding QED case. The Equation 1.3 is then modified to

$$F_{\mu\nu} = (\partial_\mu A_\nu^a - \partial_\nu A_\mu^a - ig_s [A_\mu, A_\nu]) \quad (1.8)$$

The additional non-zero commutator term in Equation 1.8 ultimately leads to interactions among the gauge fields itself. As for the QED case, massive gluons are prohibited by the gauge symmetry. The gluon-gluon interactions are in turn responsible of a rapidly increase in strength of the QCD coupling $\alpha_s = g_s^2/4\pi$ with the interaction distance. The variation of α_s with energy is shown in Figure 1.2, where natural units ($\hbar = 1$ and $c = 1$) are used, resulting in a simple inverse relation between energy and distance. At small length scales (large energies) the quarks and gluons behave as quasi-free particles because of the small coupling, so QCD can be described perturbatively. At large distances (low energies), the strong coupling is large, guaranteeing the quark confinement within subatomic particles and leading quarks and gluons to form bound states.

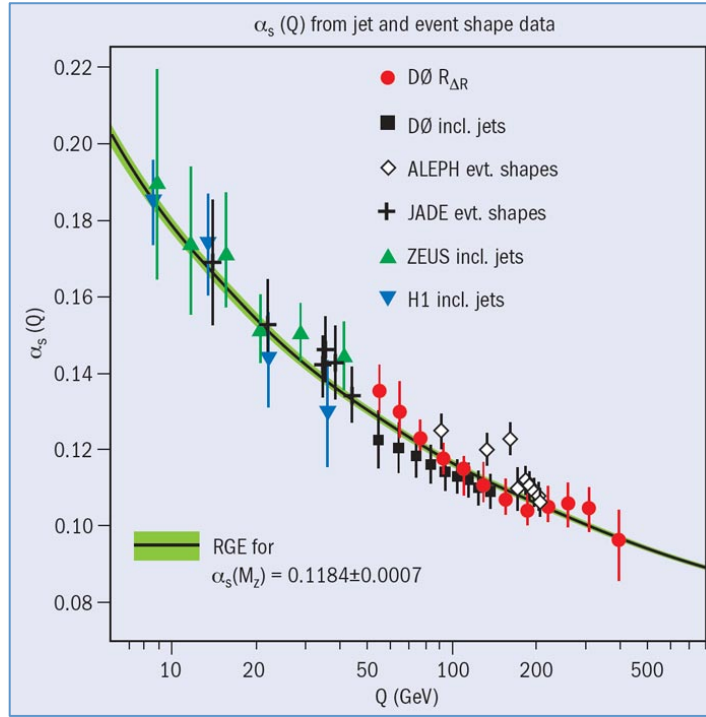


Figure 1.2 – Strong running coupling constant distribution with respect to the collision energy.

1.1.3 The Weak Interaction

Pions are the lightest hadrons, composed by a u and d quark, and form a triplet composed by the two charged pions π^\pm and the neutral one π^0 . Because of the electric charge conservation, only the π^0 can electromagnetically decay to two photons, with a lifetime of $(8.4 \pm 0.6) \cdot 10^{-17}$ s. Charged pions, instead, decay in a charged and a neutral lepton, i.e. $\pi^\pm \rightarrow \mu^\pm \nu$, with a life time of $(2,6003 \pm 0,0005) \cdot 10^{-8}$ s. This much longer lifetime has been one of the hints for the existence of a new fundamental force, weaker than the electromagnetic interaction, called indeed the weak force.

The weak interaction field Ψ is invariant under $SU(2)$ transformations and has a $V - A$ (vector – axial vector) structure that implies the weak field Ψ can be decomposed into a left-handed $\Psi_L = \frac{1}{2}(1 - \lambda_5)\Psi$ and a right-handed $\Psi_R = \frac{1}{2}(1 + \lambda_5)\Psi$ component.

A peculiarity of the weak field is to change the flavor of quarks also between different generations during interactions; however intra-generation exchange happen with a smaller probability than inter-generation ones. Interactions are only possible between up-like and down-like quarks (u/c/t-quarks \rightarrow d/s/b-quarks). An example of this is the D_0 decay: in the $D_0 \rightarrow K^- \pi^+$ process the c-quark decays to an s-quark (intra-generation) while in the less probable case $D_0 \rightarrow K^+ \pi^-$, the c-quark decays to a d-quark (inter-generation). The interaction probability is proportional to the square of the Cabibbo-Kobayashi-Maskawa (CKM) [91][92] matrix elements $|V_{qq'}|^2$.

$$V_{CKM} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \Rightarrow \quad (1.9)$$

$$\Rightarrow \begin{pmatrix} 0.97419 \pm 0.00022 & 0.2257 \pm 0.0010 & 0.00359 \pm 0.00016 \\ 0.2256 \pm 0.0010 & 0.97334 \pm 0.00023 & 0.0415^{+0.0010}_{-0.0011} \\ 0.00874^{+0.00026}_{-0.00037} & 0.0407 \pm 0.0010 & 0.99913^{+0.000044}_{-0.000043} \end{pmatrix}$$

In case the weak interaction entails the exchange of a unitary quantum of electric charge, it is considered a *charged current* interaction and its mediators are the W^\pm bosons. If no electric charge exchange occurs, the weak interaction is defined as a *neutral-current* process mediated by a Z^0 boson exchange. Weak neutral current and electromagnetic interactions interfere one each other as shown at the Large Electron-Positron Collider (LEP) experiment in the forward-background production

asymmetry found of the $e^+e^- \rightarrow \ell^+\ell^-$ process [93]. This interference suggests a sort of connection between electromagnetic and weak fields that leads to the electroweak unification.

Another striking difference between the weak interaction and the other ones is that the field quanta are massive. In the case of QED and QCD, the gauge symmetry dictates that the photon and the gluons are instead massless. Massive gauge bosons suggest that the symmetry of the weak interaction is not perfect, i.e. that the Lagrangian and the physical vacuum do not obey to the same symmetry. This is the case of a spontaneously broken symmetry mechanism which explains the W^\pm and Z^0 mass through the introduction of a scalar field.

1.1.4 Electroweak unification

The known fundamental forces differ one to each other not only in transformation properties and specific conservation rules, but also in the value of the coupling constant that determines the magnitude of the interaction mediated by that force. These constants vary with the energy involved during the interaction and for this reason they are called “running constants”. Physicists, following the successful attempt achieved by C. Maxwell in 1865 to unify electric and magnetic forces, speculate that the strong, weak, electromagnetic and gravitational forces may all be manifestations of a single universal interaction. In the late 1960s, this belief brought Weinberg, Salam and Glashow at the development of a theory including both the weak and electromagnetic interactions as a single electroweak force. This theory predicts the symmetry between electromagnetic

and weak interactions would be manifest at a very large (at that time) transferred momentum scale ($q^2 \gg 10^4 GeV^2$).

The theory postulate the existence of four massless bosons arranged in one “weak isospin” triplet ($W_\mu = W_\mu^{(1)}W_\mu^{(2)}W_\mu^{(3)}$) and one “weak hypercharge” singlet (B_μ). The field possesses the geometrical properties of a $SU(2) \times U(1)$ group. A spontaneous symmetry breaking mechanism is invoked to explain the existence of the observed massive bosons mediator of the weak force. By properly changing the reference frame and with the mediation of a scalar (Higgs) field, three massive vector bosons (W_μ^+, W_μ^- and Z_μ^0) and one massless photon A_μ appears from the massless boson combination.

$$W_\mu^\pm = \frac{1}{\sqrt{2}} [W_\mu^{(1)} \pm W_\mu^{(2)}] \quad (1.10)$$

and

$$W_\mu^{(3)} = \frac{gZ_\mu - g'A_\mu}{\sqrt{g^2 + g'^2}} \quad (1.11)$$

$$B_\mu = \frac{-g'^Z_\mu + gA_\mu}{\sqrt{g^2 + g'^2}} \quad (1.12)$$

where g and g' are the electroweak coupling constants. The resulting electroweak Lagrangian is

$$\begin{aligned} \mathcal{L}_{ew} = & \frac{g}{2} (J_\mu^- W_\mu^+ + J_\mu^+ W_\mu^-) + \frac{g}{\cos \theta_W} (J_\mu^{(3)} - \sin^2 \theta_W J_\mu^{em}) Z_\theta \\ & + g \sin^2 \theta_W J_\mu^{em} A_\mu \end{aligned} \quad (1.13)$$

where J_μ^{em} is the electromagnetic current and $J_\mu^{(3)}$ is the third component of the isospin current J_μ . J_μ^\pm are combination of the first and the second component of the isospin current J_μ :

$$J_\mu^\pm = J_\mu^{(1)} \pm i J_\mu^{(2)} \quad (1.14)$$

The electroweak coupling constants (g , g' and e) and the Weinberg angle θ_W are related as follows.

$$\begin{cases} g/g' = \tan \theta_W \\ e = g \sin \theta_W \end{cases} \quad (1.15)$$

1.1.5 Spontaneous symmetry breaking

A symmetry is said to be broken when the transformation is always invariant but for the vacuum state. The description of how to solve this “asymmetry” for the weak interaction introducing the so called Higgs mechanism follows. Consider the Lagrangian of the $SU(2)$ symmetry group describing the weak interaction.

$$\mathcal{L} = (\mathcal{D}_\mu \varphi)^\dagger (\mathcal{D}^\mu \varphi) - \mu^2 \varphi^\dagger \varphi - \lambda (\varphi^\dagger \varphi)^2 \quad (1.16)$$

where μ^2 is the mass term and λ is coupling term. The field φ is represented by a doublet of scalar complex fields

$$\varphi = \begin{pmatrix} \varphi_\alpha \\ \varphi_\beta \end{pmatrix} = \sqrt{\frac{1}{2}} \begin{pmatrix} \varphi_1 + i\varphi_2 \\ \varphi_3 + i\varphi_4 \end{pmatrix} \quad (1.17)$$

and

$$\mathfrak{D} = \partial_\mu + ig \frac{\tau_a}{2} W_\mu^a \quad (1.18)$$

where W_μ^a are three gauge fields. The weak Lagrangian considered is invariant under $SU(2)$ phase transformation

$$\varphi \rightarrow \varphi' = e^{\frac{i\alpha_a(x)\tau_a}{2}} \varphi \quad (1.19)$$

and the following gauge transformation

$$\varphi(x) \rightarrow \varphi'(x) = \left(1 + \frac{i\alpha(x)\tau}{2} \right) \varphi(x) \quad (1.20)$$

$$\mathbf{W}_\mu \rightarrow \mathbf{W}_\mu - \frac{1}{g} \partial_\mu \boldsymbol{\alpha} - \boldsymbol{\alpha} \times \mathbf{W}_\mu$$

It is possible to find a solution of the form

$$V(\varphi) = \mu^2 \varphi^+ \varphi + \lambda (\varphi^+ \varphi)^2 \quad (1.21)$$

In both cases with $\mu^2 > 0$ and $\mu^2 < 0$. The first degenerate solution leads to a system of four scalar massive particles interacting via three massless gauge bosons. The second and more interesting case has a manifold of potential minimum degeneracy in energy represented by

$$\varphi^+ \varphi = \frac{1}{2}(\varphi_1^2 + \varphi_2^2 + \varphi_3^2 + \varphi_4^2) = -\frac{\mu^2}{2\lambda} \quad (1.22)$$

with the manifest $SU(2)$ symmetry. This solution can be rewritten as

$$\begin{aligned} \varphi_1 = \varphi_2 = \varphi_4 &= 0 \\ \varphi_3^2 &= -\frac{\mu^2}{\lambda} = v^2 \end{aligned} \quad (1.23)$$

where the symmetry is hidden. The corresponding vacuum state is

$$\varphi_0 = \sqrt{\frac{1}{2}} \begin{pmatrix} 0 \\ v \end{pmatrix} \quad (1.24)$$

The expansion of the ground state gives the field

$$\varphi(x) = \sqrt{\frac{1}{2}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix} \quad (1.25)$$

where $h(x)$ is the Higgs scalar field. Inserting this result in the Lagrangian of the Equation 1.10 the system described is composed by three massive vector bosons (the W^\pm and Z^0 mediators) and by a fourth scalar particle, the Higgs boson. This mechanism can be extended to the $U(1) \times SU(2)$ electroweak field to complete the SM scheme.

1.2 Top Quark

The tau lepton was the first particle of third generation to be discovered [3], in 1975. Only a few years later, in 1977, the evidence of the Y meson found at Fermilab[4] and interpreted as a $b\bar{b}$ bound state, led to the first evidence of a third generation quark, the bottom one. With this discovery the SM theory needs the existence of an additional quark, the top quark, as the third-generation partner of the bottom quark. Indeed, the existence of such a third generation quark doublet, in conjunction with the presence of three lepton generations, cancels some anomaly divergences of the theory.

In 1995 the top quark has been finally discovered by the CDF and D0 collaborations [5][6] at the $p\bar{p}$ collider Tevatron; its mass $m_t = (173.5 \pm 0.6(stat.) \pm 0.8(sist.)GeV$ [7] is the heaviest of all fundamental particles known, about 40 times heavier than the b quark and comparable to the mass of an atom of Rhenium ($Z = 75$). Due to its very short lifetime ($\tau_t \sim 0.5 \cdot 10^{-24} s$), the top quark decays before to hadronize ($t_{had} \sim 3 \cdot 10^{-24} s$) and offers a unique opportunity to study the properties of a bare quark, including polarization effects that can be studied from the angular correlations between the decay products.

The large value of m_t also implies a large coupling to the Higgs boson, making the top quark physics an important channel to study Higgs related constraints and to identify associated production channel as ttH . In addition it is one of the main backgrounds in many Higgs related analyses. The top quark plays also an important role in many beyond SM physics scenarios, as described in the following paragraphs.

1.2.1 Top quark decay

The extremely heavy top quark has the very short lifetime of approximately 10^{-24} seconds, which is short enough to decay before to hadronize [16]. The top quark is unique in this respect. No free quark has ever been found so the top quark supplies the only probe of the behavior of bare quarks.

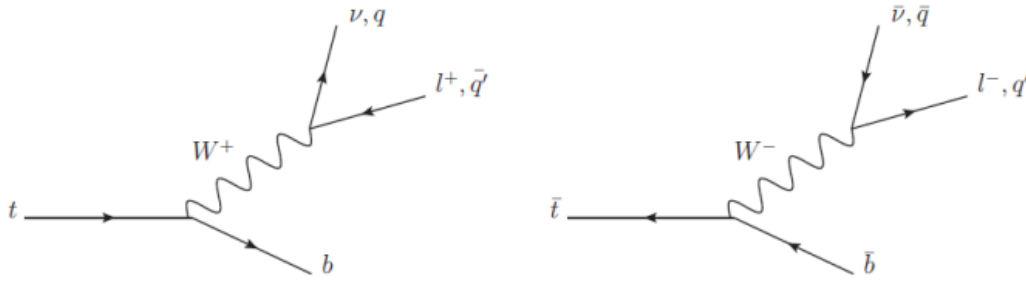


Figure 1.3 – Top and anti-top decay chains.

Due to the top quark extremely short lifetime, detectors can only measure its decay products. The top quark decay in a b quark and a W boson that can in turn decay in a lepton plus neutrino or in a light quark pair; the Feynman diagrams of these processes are shown in Figure 1.3, are dominated by the $t \rightarrow Wb$ process since the CKM

element V_{tb} is close to 1. The produced b quark hadronize to a B meson or baryon that decays, via $b \rightarrow Xc$ or $b \rightarrow Xu$ processes. According to CKM matrix, these decays are preferred, but both have a reduced ratio because of intra-generation suppression in weak interactions, briefly discussed in Paragraph 1.1.3. This contributes to the relatively long lifetimes of the B hadrons, most of them have a decay time of approximately 1.5 ps [16], that is exploited as the principal discriminator factor in b quark identification algorithms.

The $W^+(W^-)$ daughter of the top(anti-top) decays either leptonically into an anti-lepton(lepton) and its associated neutrino(anti-neutrino) or hadronically into a qq' pair. The $t\bar{t}$ events are classified according to the decays of the two W^\pm bosons. When both W^\pm bosons decay hadronically the event is called a “fully hadronic” event ($BR \approx 0.46$). When both W^\pm bosons decay leptonically the event is defined as a “dilepton” event ($BR \approx 0.09$). When one W^\pm decays leptonically and the other hadronically the event is a “lepton + jets” event ($BR \approx 0.45$). This scheme is summarized in Figure 1.4.

1.2.2 Top Quark pair production

In proton-proton (pp) high energy collisions, like at LHC, the top quark dominant production mechanism is the top/anti-top pair ($t\bar{t}$) production, a strong interaction process whose leading order diagrams coming from gluon-gluon (gg) and quark-antiquark ($q\bar{q}$) interactions are shown in Figure 1.5 while next to leading order (NLO)

mixed quark–gluon (qg), or antiquark–gluon ($\bar{q}g$), processes are shown in diagrams in Figure 1.6.

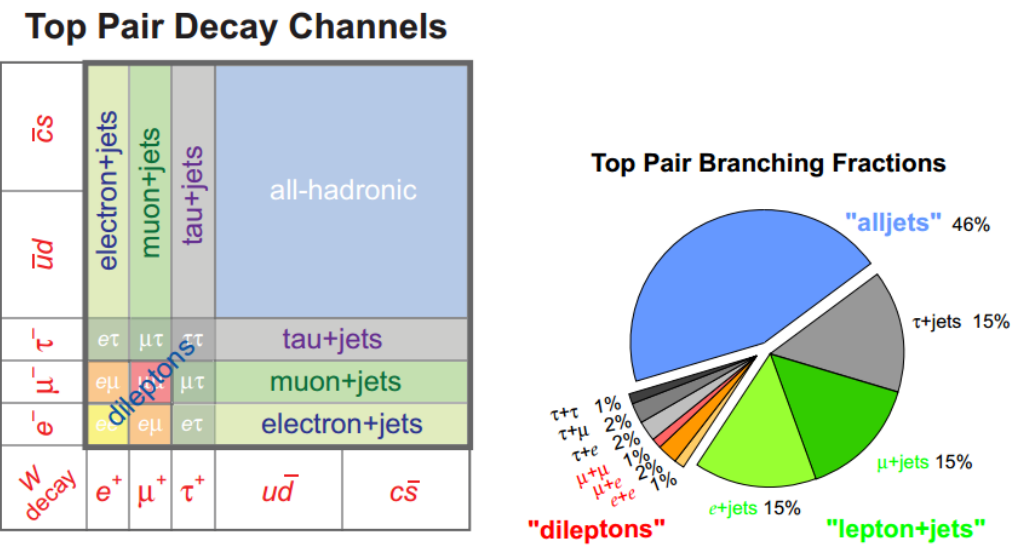


Figure 1.4 – $t\bar{t}$ pair decay channels (left) and branching ratios (right).

To quantify the $t\bar{t}$ production, as for any other particle production, it is necessary to measure the cross section of the process. It expresses the likelihood of particle interactions. In classical mechanics the cross section is related to the effective area of collision of two, or more, bodies.

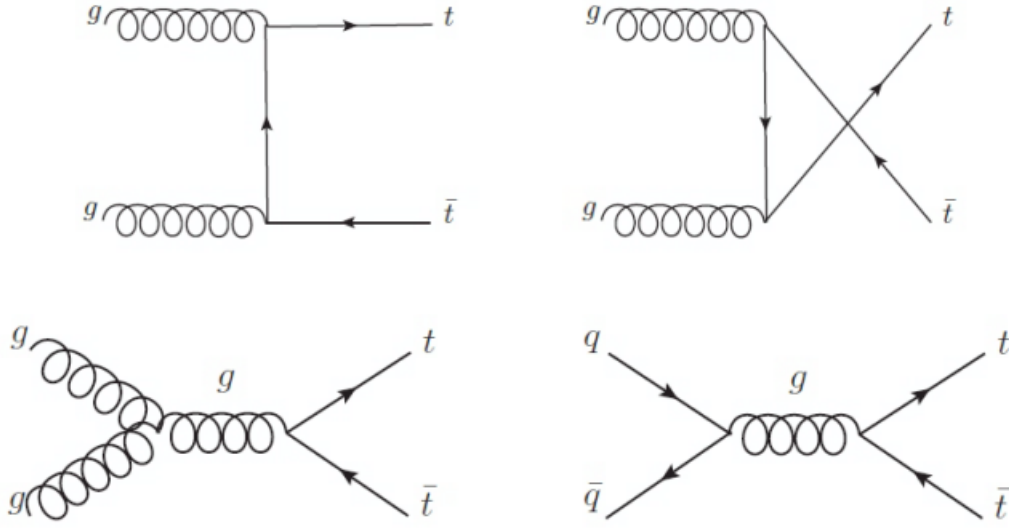


Figure 1.5 – Leading order Feynman diagrams for top/anti-top pair production via gluon fusion and quark/anti-quark annihilation processes.

In particle physics the cross section is a more complicated quantity including wavelength probability terms from quantum mechanics, but it is still proportional to the interaction probability. In a high energy colliders such as the LHC, the protons can scatter inelastically producing new particles, as in the $pp \rightarrow t\bar{t}$ case, in addition to the elastic $pp \rightarrow pp$ scattering; all those possible scattering processes are considered in the total inclusive cross section. The exclusive cross section for a process can be thought of as the probability for that process to happen.

The general cross section formula for a process is given by:

$$\sigma = \frac{N_{events}}{\varepsilon \int \mathcal{L} dt} \quad (1.26)$$

where N_{events} is the number of observed events, ε is the efficiency of the detector and of the particle selection cuts that are applied; $\int \mathcal{L} dt$ is the integrated luminosity, meaning

the luminosity obtained during the data acquisition; the instantaneous luminosity is a characteristic of the collider given by the relation

$$\mathcal{L} = f \frac{n_1 n_2}{4\pi\sigma_x\sigma_y} \quad (1.27)$$

where f is the collision frequency ($f_{LHC} \approx 11 \text{ kHz}$), n_1 and n_2 are the numbers of particles contained in each bunch ($n_{LHC} \approx 10^{11}$) and σ_x and σ_y are the beam particle distribution along two orthogonal axis with respect to the beam direction ($\sigma_{LHC} \approx 50 \mu\text{m}$). The cross section is usually measured in barn (b) and its multiples and submultiples; one barn corresponds to 10^{-24} cm^2 .

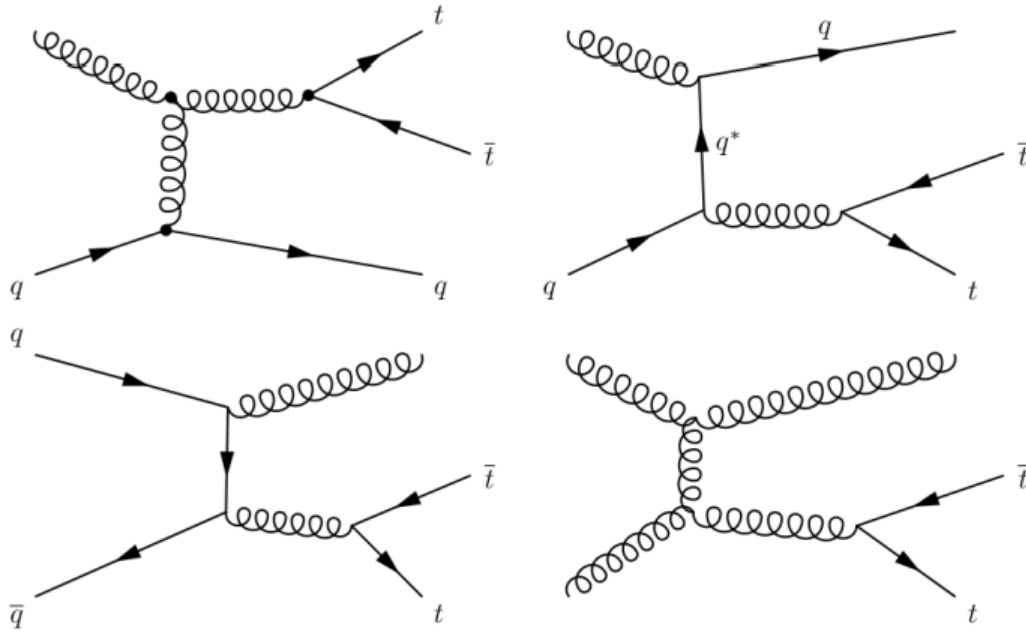


Figure 1.6 –Examples of next to leading order Feynman diagrams for top/anti-top pair production via gluon fusion and quark/anti-quark annihilation processes.

At the energies of LHC, high enough to consider quarks and gluons as quasi-free particles, pp collisions can be described in terms of interactions between their constituent partons (i.e. quarks and gluons); the cross section is then convoluted with the Parton Distribution Function (PDF) of the proton in order to deal with the proton internal structure. The PDFs give the momentum distribution of the quarks, anti-quarks and gluons inside the proton. They are determined from the combination of a large amount of experimental data on deep-inelastic scattering (especially from HERA) and from jet and heavy quark production results at hadron colliders.

In Figure 1.7 the quarks and gluons momentum distribution from PDF is shown for the proton, where x is the parton momentum fraction. At LHC $\sqrt{s} = 7(14)TeV$ energies, around the 80 (90)% of the total cross section is due to the gg contribution, while the remaining is mostly due to $q\bar{q}$ interactions. At the Tevatron $p\bar{p}$ collider, the situation was reversed with the $q\bar{q}$ contribution dominating.

The inclusive production cross section of the process $pp \rightarrow t\bar{t}$ can be expressed using the factorization theorem as a convolution of parton distribution functions (PDF) and the partonic cross section

$$\sigma_{pp \rightarrow t\bar{t}}(s, m_t) = \sum_{i,j=q,\bar{q},g} \int dx_i dx_j f_i(x_i, \mu_f^2) f_j(x_j, \mu_f^2) \cdot \hat{\sigma}_{ij \rightarrow t\bar{t}}(\hat{s}, m_t, \mu_f, \mu_r, \alpha_s) \quad (1.28)$$

The sum runs over all quarks and gluons contribution while x_i are the parton momentum fractions with respect to the proton momenta, $f_i(x_i, \mu_f^2)$ are the proton PDF, $\mu_{f(r)}$ are the factorization and regularization scales, α_s is the strong coupling and

$\hat{s} \equiv (x_i x_j s)$ is the partonic center-of-mass energy. The cross section strongly depends on top mass m_t and the squared center of mass energy of the collider $s = 4E_{beam}^2$. The dependence from the regularization scale arises from the fact that the partonic cross section is evaluated at a fixed perturbation order, neglecting higher order contributions. Such dependences become weaker and weaker as we add higher order corrections to calculation. The factorization scale, on the other hand, indicates the transition between the perturbative regime, which belongs to the partonic cross section, and the non-perturbative one, included in the PDF definition. Of course the physical cross section should not depend on the two scales mentioned above, as they are not physical parameters. The dependence arise from the unavailable approximations used in the calculation and the uncertainties on the scales is one of the major sources of uncertainty for the cross section prediction. Usually the renormalization and factorization scales are set equal to the mass of the top

$$\mu = \mu_r = \mu_f = m_t . \quad (1.29)$$

In some cases, for example for differential cross sections studies, these scale factors may assume other values, for example the transverse momentum of a jet ($p_{T,jet}$) or the top quark pair invariant mass ($m_{t\bar{t}}$). In order to estimate the uncertainty coming from this arbitrary choice, the scale factors are changed within a certain range, often $[\mu/2, 2\mu]$ is used.

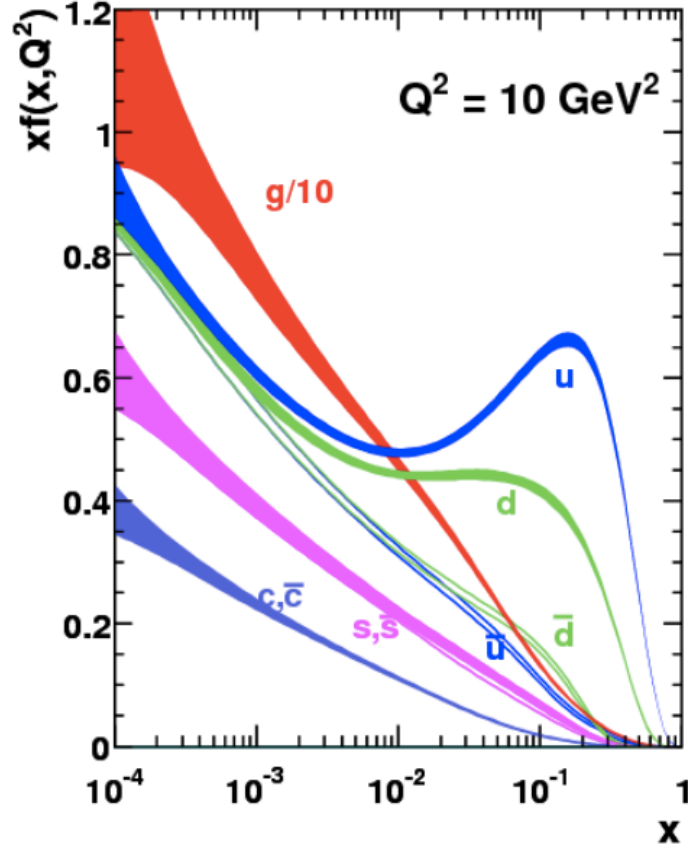


Figure 1.7 – MSTW2008NLO PDF set [94] prediction for quarks and gluons momentum distribution inside the proton. Tevatron working point is about $x \sim 0.1$ while for LHC it is ~ 0.02 .

1.2.3 $t\bar{t}$ pair cross section studies

The $t\bar{t}$ cross section has been experimentally measured for the first time at Tevatron by the CDF[8] and D0[9] collaborations; the result for a top quark mass of $m_{top} = 173 \text{ GeV}$ is

$$\sigma_{t\bar{t}}^{Tevatron}(p\bar{p}, 1.96 \text{ TeV}) = 7.08 \pm 0.36 \text{ pb} \quad (1.30)$$

In Figure 1.8 the experimental cross section is compared with the exact NLO and the approximate NNLO predictions showing the last one is in better agreement with data.

The theoretical $t\bar{t}$ production cross section at the higher energies of LHC is far greater than the one at Tevatron because of its strong dependence of the collision energy.

The expected values for the operating LHC energies up to now are:

$$\begin{aligned}\sigma_{t\bar{t}}^{theory}(7\text{ TeV}) &= 172.0^{+0.9}_{-1.2}\text{ pb} \\ \sigma_{t\bar{t}}^{theory}(8\text{ TeV}) &= 238^{+22}_{-24}\text{ pb}\end{aligned}\tag{1.31}$$

The ATLAS[10] and CMS[11] collaborations measured the $t\bar{t}$ cross section in pp collisions at a center of mass energy of 7 TeV observing:

$$\sigma_{t\bar{t}}^{ATLAS}(7\text{ TeV}) = 177^{+8}_{-7}(\text{sys}) \pm 3(\text{stat}) \pm 7(\text{lumi})\text{ pb}\tag{1.32}$$

$$\sigma_{t\bar{t}}^{CMS}(7\text{ TeV}) = 165 \pm 10.6(\text{sys}) \pm 2.2(\text{stat}) \pm 7.8(\text{lumi})\text{ pb}\tag{1.33}$$

in good agreement with the theoretical predictions. These experimental results, obtained from the combination of measurements performed in different decay channels, are in good agreement with the predicted value as can be seen from plots in Figure 1.9, where results are compared with the NLO and NNLO predictions respectively. Once again the NNLO prediction better match the experimental measurements. All the recent ATLAS and CMS results for each channel considered are shown in Figure 1.10. In Figure 1.11 both NNLO theoretical $t\bar{t}$ cross section distribution for pp and $p\bar{p}$ interactions versus the

center of mass energy are shown together with the Tevatron and LHC experimental results. In both cases the experimental measurements are in good agreement with theory.

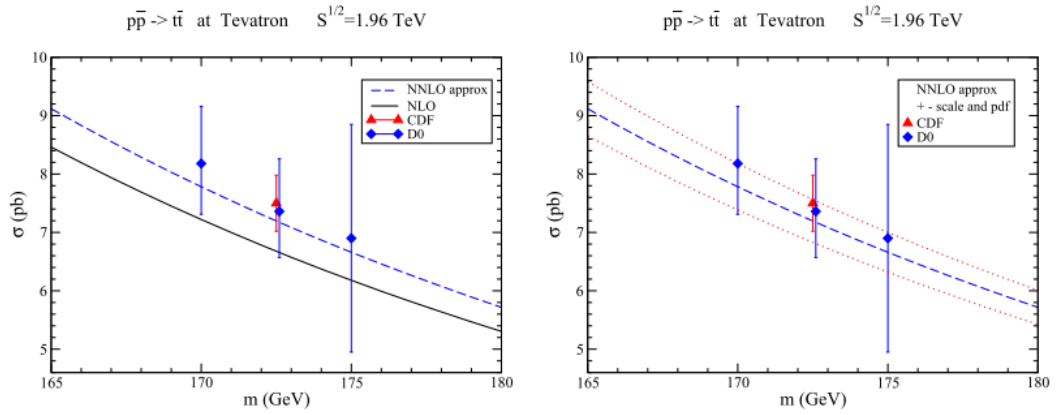


Figure 1.8 – Tevatron $t\bar{t}$ cross section measurements compared, on the left plot, with NLO and NNLO prediction and, on the right, with NNLO scale variation uncertainties estimated varying PDF scale ($\mu = m_t$) of a factor two [8][9].

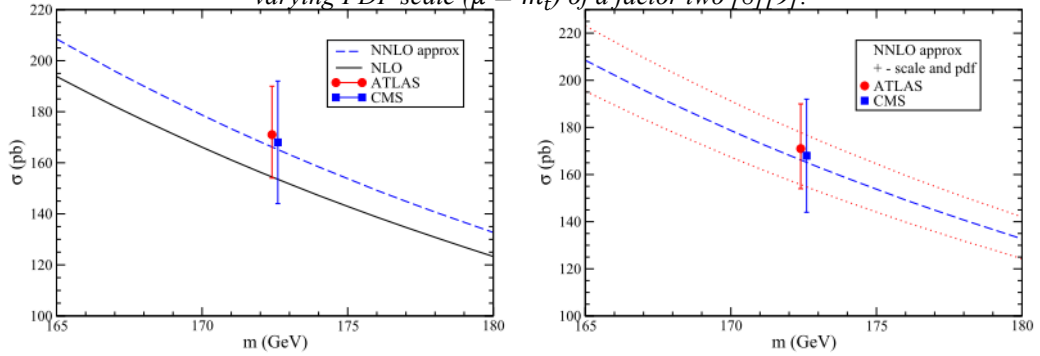


Figure 1.9 – LHC $t\bar{t}$ cross section measurements compared, on the left plot, with NLO and NNLO prediction and, on the right, with NNLO scale variation uncertainties estimated varying PDF scale ($\mu = m_t$) of a factor two [10][11].

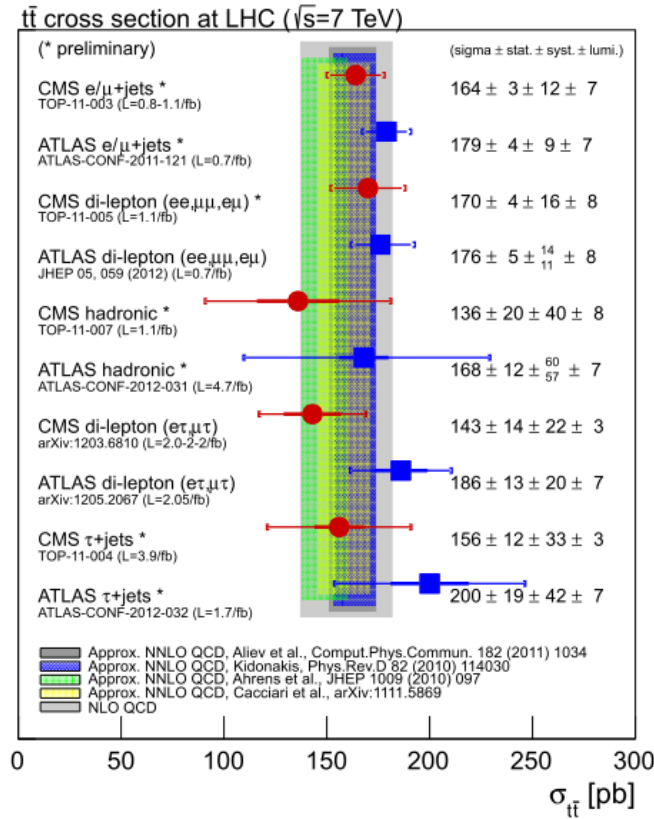


Figure 1.10 – Summary of the most precise measurements of $\sigma_{t\bar{t}}$ per decay mode and experiment, compared with several theory predictions at NLO and approximate NNLO QCD.

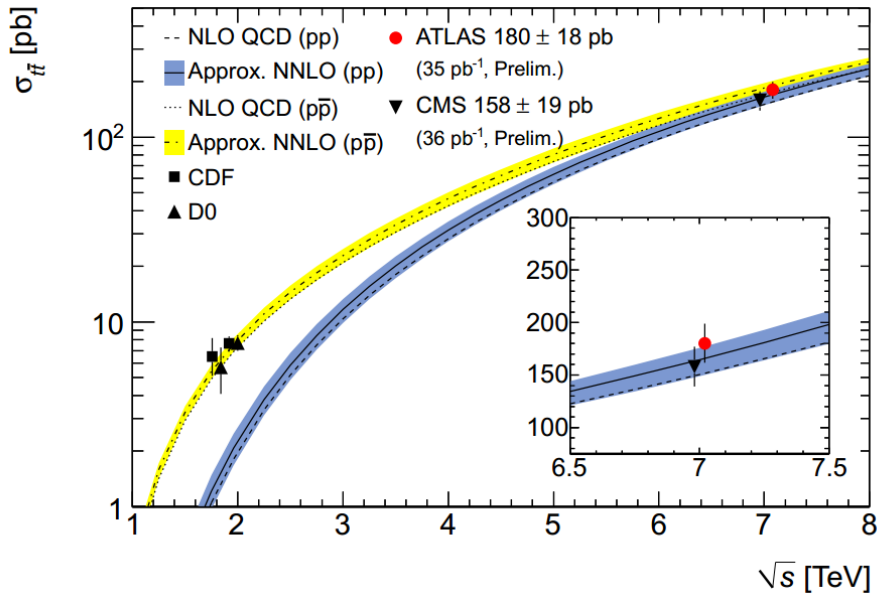


Figure 1.11 – Inclusive top/anti-top production cross section predicted for LHC and compared with ATLAS and CMS measurements.

The $\sigma_{t\bar{t}}$ measurement has been also performed at a center of mass energy of 8 TeV by both ATLAS[24] ($\int \mathcal{L} dt = 5.8 fb^{-1}$) and CMS[25] ($\int \mathcal{L} dt = 2.4 fb^{-1}$) in the *lepton+jets* channel founding results compatible with the theoretical expectations.

$$\sigma_{t\bar{t}}^{ATLAS}(8TeV) = 241 \pm 31 (sys) \pm 2(stat) \pm 9(lumi) pb \quad (1.34)$$

$$\sigma_{t\bar{t}}^{CMS}(8TeV) = 228_{-26}^{+29} (sys) \pm 9(stat) \pm 10(lumi) pb \quad (1.35)$$

1.2.4 Differential $t\bar{t}$ cross section

The measurement of the differential $t\bar{t}$ production cross section constitutes a very important test of the standard model predictions and a sensitive channel for the presence of new physics. Prominent is the role of the $t\bar{t}$ invariant mass distribution, which may be significantly modified by the presence of resonances that decay into top pairs. Theoretical predictions for such distributions exist both in fixed order QCD and in SCET model[84] showing theoretical errors between 10% and 15%, depending on the $m_{t\bar{t}}$ value assumed. Similar importance has the cross section dependence from the p_T of the top quark. The corresponding theoretical prediction [85] for the LHC data taking at $\sqrt{s} = 14TeV$ is shown in Figure 1.12. The *lepton+jets* final state channel is probably the most promising among the $t\bar{t}$ decaying channels for its high efficiency given by the leptonic trigger and the not so high E_T^{miss} impact thanks to the hadronic decaying top quark. For these reasons here and in the rest of the thesis, the attention will be mainly focused on the *lepton+jets* channel. The differential cross section may be calculated

either after extrapolation to the full phase space, at the level of partons before hadronization (*parton level*), or only within a reduced, or fiducial, phase space considering only objects visible by the detector (*particle level*); in this last case only the detector response correction is used. The first definition may be needed in order to compare with fixed order QCD calculations, while the second definition is closer to what is measured experimentally and can be easier compared with MC simulations.

The first attempt of differential cross section evaluation has been done by the Tevatron experiments resulting in the measurements of the cross section as a function of the transverse momentum of the top quark by DØ [86] with an integrated luminosity of $\mathcal{L} = 1fb^{-1}$ and as a function of the invariant mass of the $t\bar{t}$ system by CDF [87] with an integrated luminosity of $\mathcal{L} = 2.7fb^{-1}$. Both results are consistent with the standard model cross section predictions as shown in Figure 1.13.

Thanks to the large abundance of top quark pair production due to the high cross section, at the LHC collider differential cross section measurements can be performed with increasingly precision as function of several kinematic variables. This improves the reliability of the measure and widens the horizon for new physic searches. Several measurements with increasing statistic have been performed by the ATLAS and CMS collaboration on different decay channels.

Results presented in this thesis exactly deals with one of the most recent and precise differential cross section measurements performed at ATLAS. These results in the $lepton+jets$ obtained with a 2011 data statistic corresponding to an integrated luminosity $\mathcal{L} = 4.7fb^{-1}$ and collected with a center of mass energy $\sqrt{s} = 7TeV$, will be detailed described in the following chapters together with the analysis method used. In this

paragraph is instead reported the differential cross section results, already published in a paper [89], evaluated in the *lepton+jets* channel by the previous analysis working group (Figure 1.14). An integrated luminosity of $\mathcal{L} = 2.6 \text{ fb}^{-1}$ has been used in this case. All differential cross section results presented, are evaluated at the *parton level*.

An ongoing analysis on the data collected during 2012 is also presented in this thesis; this analysis is specifically looking at very energetic phase space regions where the standard model predictions are mostly pushed forward. In this last case the differential cross section will be evaluated at the *particle level*.

The CMS collaboration most recent results (at *parton level*) in the *lepton+jets* and *dilepton* channels as been performed using the 2011 data corresponding to an integrated luminosity of $\mathcal{L} = 5.0 \text{ fb}^{-1}$ and $\sqrt{s} = 7 \text{ TeV}$ [88]. A good agreement has been found with all the predictions considered as shown by the *lepton+jets* results reported in Figure 1.15.

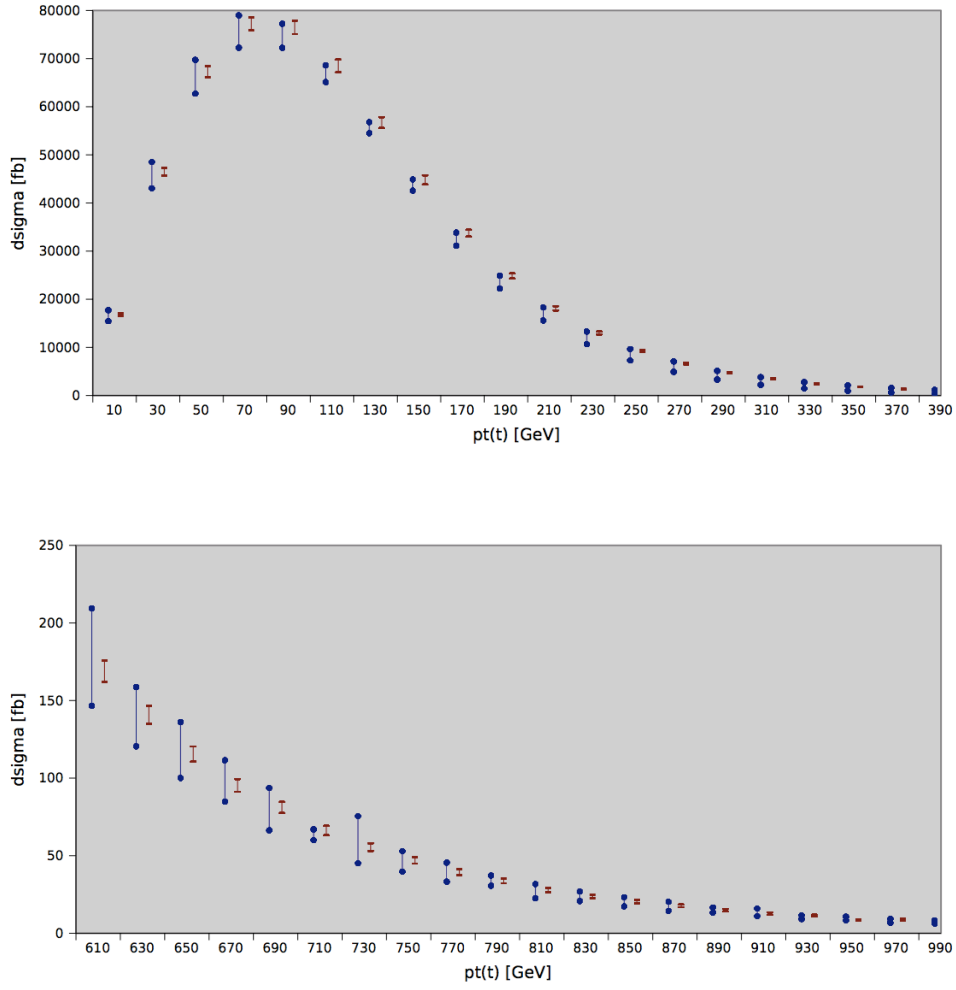


Figure 1.12 – NLO QCD predictions [90] for the transverse momentum of the top quark at the 14 TeV LHC. Blue error bars correspond to the central MSTW pdf set and scale variation by a factor of two around $\mu = m_t$. Dark red error bands correspond to 1 standard deviation of MSTW pdf error sets for fixed renormalization and factorization scale at $\mu = m_t$. Note that the red and blue bars can be off-set because at NLO the central scale does not necessarily corresponds to the center of the blue bar. In this case, it seems that it is towards the upper value of the blue bar.

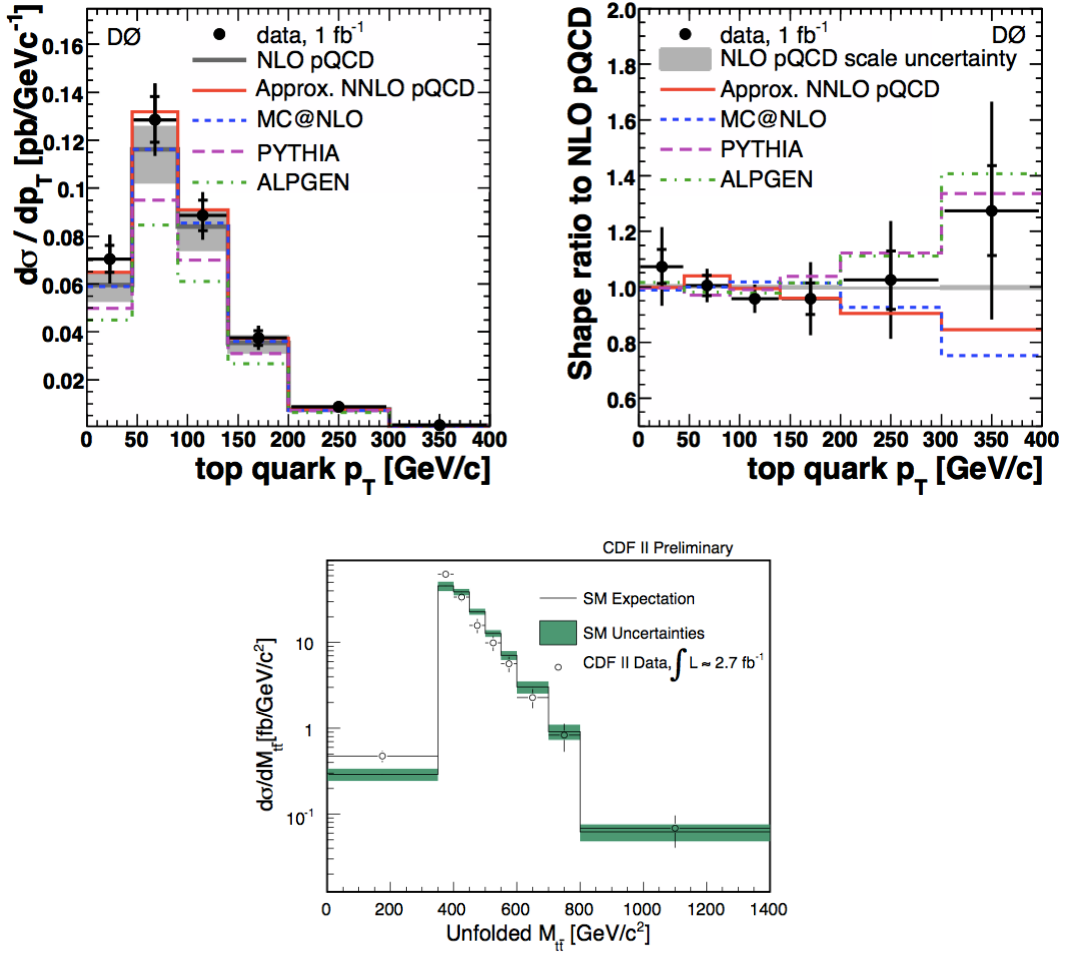


Figure 1.13 – Differential cross section (top-left) and relative ratio (top-right) data (points) obtained by the D0 experiment as a function of top-quark p_T (two entries per event) [86] compared with expectations from NLO (solid lines), from an approximate NNLO calculation, and for several event generators (dashed and dotted lines). In the bottom plot it is shown the differential $t\bar{t}$ cross section (circles) obtained by the CDF collaboration as a function of $m_{t\bar{t}}$ [87] compared to the standard model expectation (line).

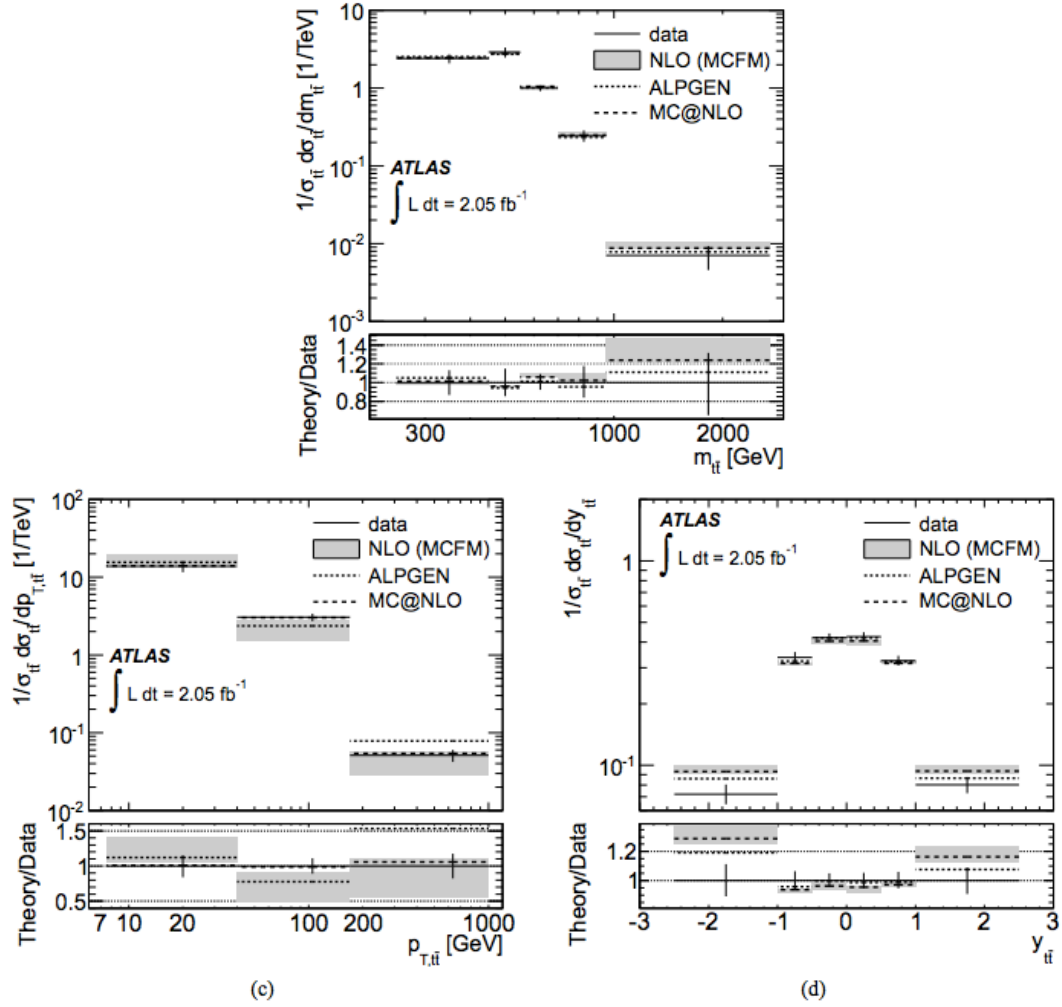


Figure 1.14 – Electron and muon channels normalised differential $t\bar{t}$ production cross section obtained by the ATLAS collaboration as a function of the $m_{t\bar{t}}$ (top), $p_{T,t\bar{t}}$ (bottom-left) and $y_{t\bar{t}}$ (bottom-right). The inner (outer) error bars indicate the statistical (combined statistical and systematic) uncertainty. The measurements is compared to the NLO prediction from MCFM. [89]

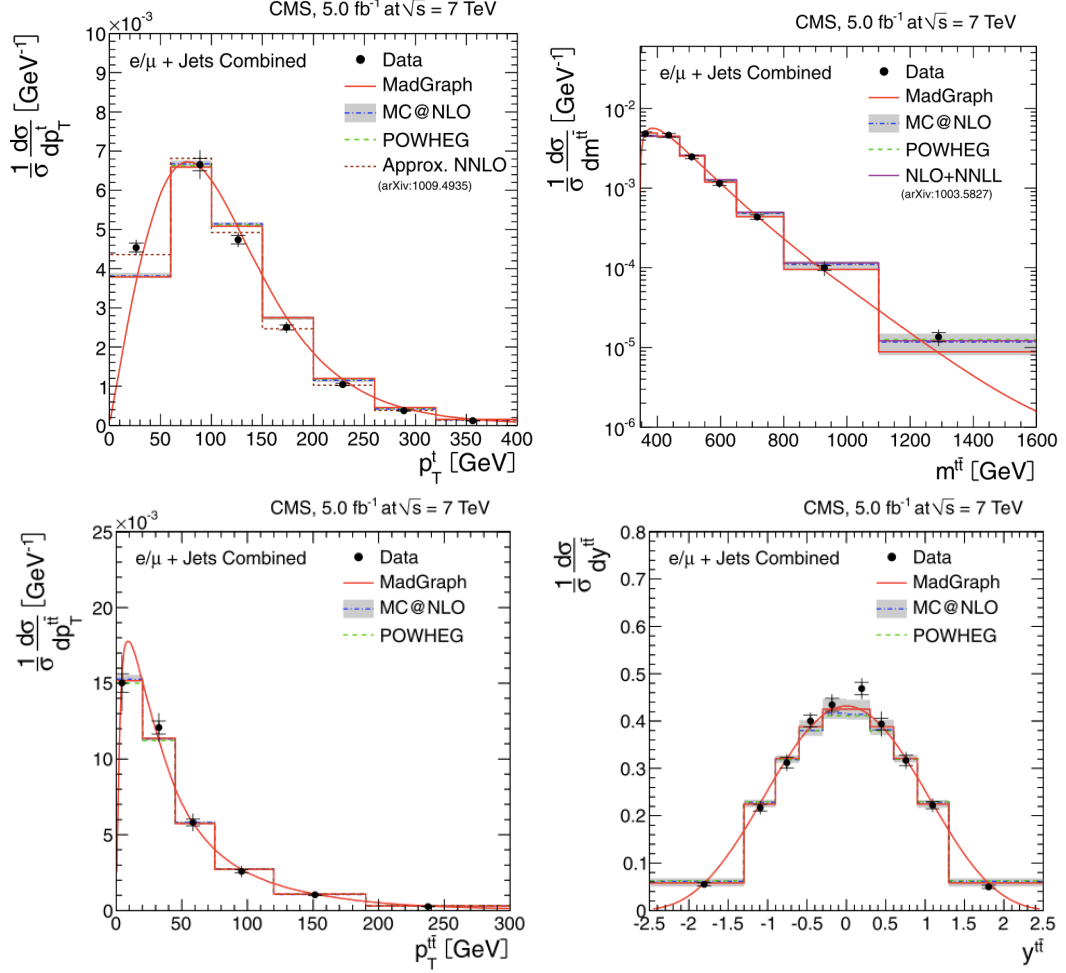


Figure 1.15 – Electron and muon channels normalised differential $t\bar{t}$ production cross section obtained by the CMS collaboration as a function of the p_T^t (top-left), $m_{t\bar{t}}$ (top-right), $p_T^{t\bar{t}}$ (bottom-left) and $y_{t\bar{t}}$ (bottom-right). The inner (outer) error bars indicate the statistical (combined statistical and systematic) uncertainty. The measurements are compared to predictions from MADGRAPH, POWHEG, and MC@NLO. The MADGRAPH prediction is shown both as a curve and as a binned histogram. [88]

1.2.5 Single top quark production

Top quarks can also be produced as single quarks in electroweak interactions. Three different mechanisms contribute to single top quark production and their leading-order Feynman diagrams are the *t-channel*, the *s-channel* and the *Wt-channel* (shown in Figure 1.16).

The first observation of single top quark production was achieved by the Tevatron experiments CDF and D0 in 2009 [12][13]. It constituted a very remarkable result especially considering the extremely low cross section of the process at the Tevatron energy. This is not anymore true at LHC where the higher energies reached lead to a significative contribution from single top production. The expected cross sections at $\sqrt{s} = 7 \text{ TeV}$ and $\sqrt{s} = 8 \text{ TeV}$ are shown in Table 1.1.

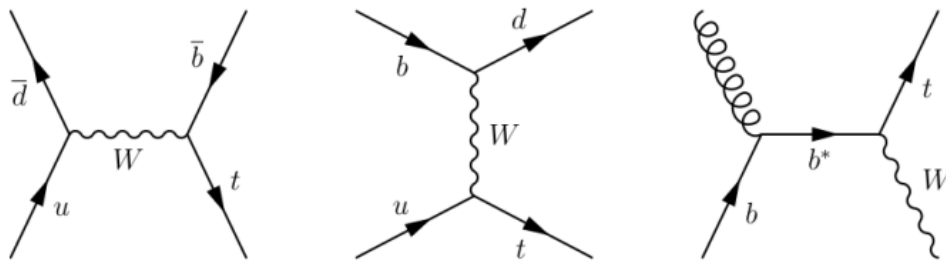


Figure 1.16 – Single top production channels at leading order. From left to right *s-channel*, *t-channel* and *Wt-channel* are shown.

Channel	7 TeV [pb]	8 TeV [pb]
t-channel	$64.6^{+2.7}_{-2.0}$	$87.8^{+3.4}_{-1.9}$
s-channel	4.6 ± 0.2	5.6 ± 0.2
Wt	15.7 ± 1.1	22.4 ± 1.5

Table 1.1 – Expected single top quark production cross sections in different channels at a center-of-mass energy of 7 TeV, given by approximate NNLO [21][22][23].

1.2.6 Top quark mass

The top quark mass m_t is a free parameter of the Standard Model and must be determined experimentally. A precise determination of m_t is important because quantum loops involving top quarks induce large corrections to theory predictions for many precision electroweak observables, including the mass of the Higgs boson.

Presently, the most precise direct measurement of the top quark mass has been done by the Tevatron experiments[7]

$$m_t = (173.5 \pm 0.6(stat.) \pm 0.8(sist.) GeV \quad (1.36)$$

with a precision of 0.6% that makes the mass of the top quark not only the highest but also the one known with better precision among all the quarks. In Figure 1.17 the most recent direct m_t results are summarized. Indirect constraints on m_t can be obtained from precision measurements of the parameters of the electroweak theory. The mass of the W boson can be expressed as

$$m_W^2 = \frac{\pi\alpha(M_Z^2)/\sqrt{2}G_F}{\sin^2\theta_W \cdot (1 - \delta_r)} \quad (1.37)$$

where $\alpha(M_Z^2)$ is the electromagnetic coupling constant, G_F is the Fermi constant and θ_W is the electroweak mixing angle ($\sin^2\theta_W = 1 - \frac{m_W^2}{m_Z^2}$). The term δ_r contains contributions from higher order electroweak loop diagrams involving the top quark, which depend quadratically on m_t .

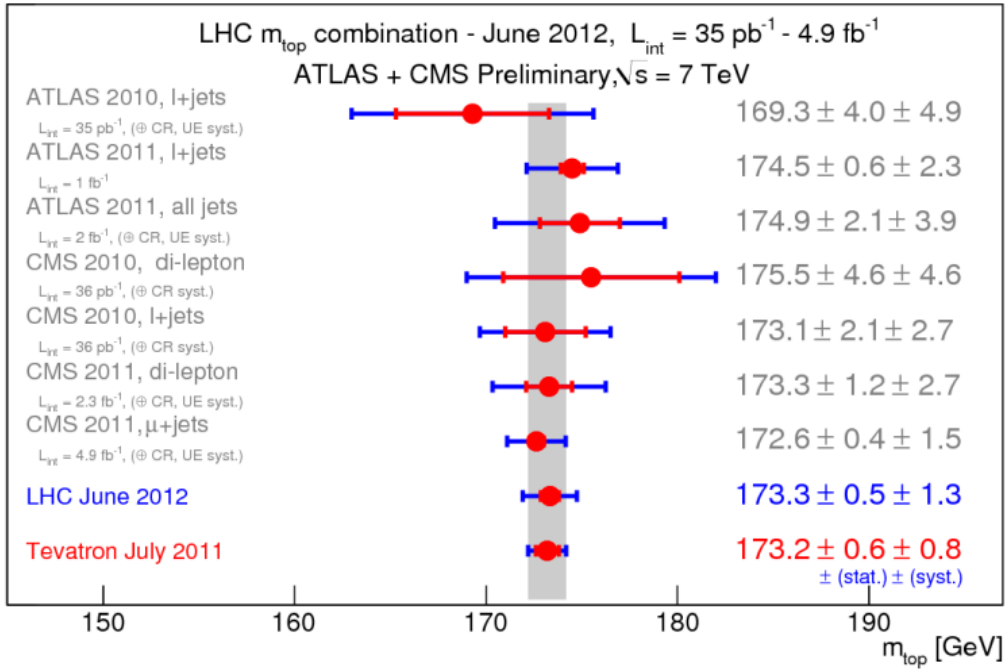


Figure 1.17 – Overview of the top quark mass measurements from ATLAS and CMS, including the latest CDF and D0 combination.

The most recent indirect constraint on m_t based on electroweak precision measurements [18] is

$$m_t^{indirect} = 178.9^{+11.7}_{-8.6} GeV \quad (1.38)$$

which is in good agreement, even if with a much larger uncertainty, with the direct measurements.

Since the δ_r term depends logarithmically on the Higgs mass, it is possible to indirect constraints m_H from global electroweak fits including direct measurements of m_t . The most recent determinations of those constraints fits with the Higgs boson discovery accomplished at LHC by the ATLAS[19] and CMS[20] experiments.

Chapter 2

The Large Hadron Collider and the ATLAS Detector

2.1 LHC

The LHC accelerator is a proton-proton collider located at the CERN laboratories, in the same 27 km long tunnel used for the LEP collider. As shown in Figure 2.1, this tunnel is situated ~ 100 m under the Geneva periphery. LHC is designed to accelerate protons up to an energy of 7 TeV producing collision at a center-of-mass energy of 14 TeV at a maximum luminosity of about $10^{34} \text{ cm}^{-2}\text{s}^{-1}$. These high performances will be reached only after a quite long period of time. The LHC started its operations in 2008; during the 2010 and 2011 runs LHC has been operated at a center of mass energy of 7

TeV. In the same year the luminosity was gradually pumped up to a maximum value of about $\mathcal{L} = 4 \cdot 10^{33} \text{ cm}^{-2}\text{s}^{-1}$. In the 2012, the center-of-mass energy has been instead increased to 8 TeV.

The two beams circulating in the accelerator are structured in 3564 bunches (BC) separated by 25 ns each. Presently only 2808 were filled with a bunch separation of 50 ns. In order to keep beams onto their circular trajectory 1232 superconducting dipole magnets (at a current of 11.85 kA and a temperature of 1.9 °K) generate a magnetic field of 8.4 T (see Figure 2.2). The focusing system consists of 392 superconducting magnets quadrupoles producing a 6.8 T field. Beams circulate in opposite directions into two separate ultrahigh vacuum chambers at a pressure of 10^{-10} torr.

Before entering into the LHC beam-pipe, proton are pre-accelerated up to an energy of 300 GeV by an injection chain composed by a linear accelerator, the **Linac2**, and three synchrotrons, Proton Synchrotron Booster (**PSB**), Proton Synchrotron (**PS**) and Super Proton Synchrotron (**SPS**). In LHC the beams are further accelerated by 16 radiofrequency cavities with a maximum electric field of 5.5 MV/m.

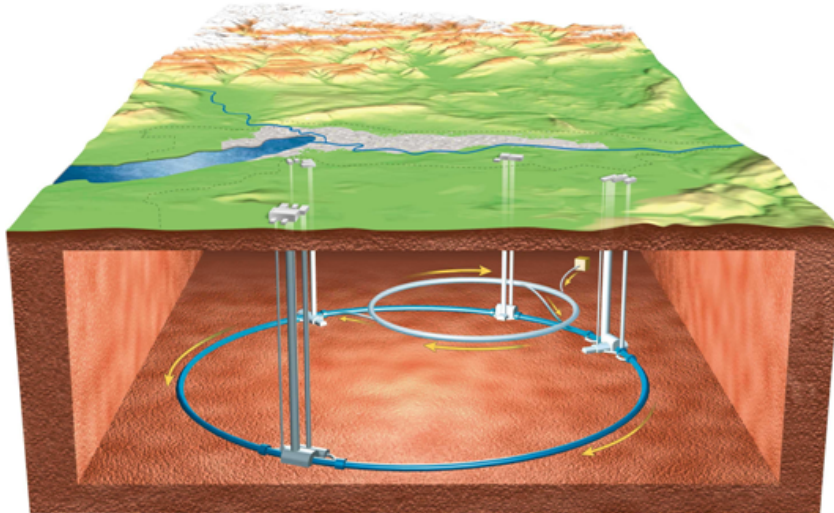


Figure 2.1 – The LHC apparatus located in the underground near Geneva

Four interaction regions have been instrumented along the tunnel and host the following experiments:

- **ATLAS (A Toroidal Lhc ApparatuS)** is one of the two multipurpose experiment which will work mainly at high luminosity to discover the Higgs boson and eventually signatures of new physics.
- **CMS (Compact Muon Spectrometer)** is the second multipurpose experiment that pursues the same physics goals as ATLAS using different and complementary technologies. It will work mainly at a high luminosity.
- **LHCb** will perform accurate measurements in the flavour physics of the B meson (CP violation).
- **ALICE (A Large Ion Collider Experiment)** is dedicated to the study of the quark-gluon plasma in heavy nucleus collision.

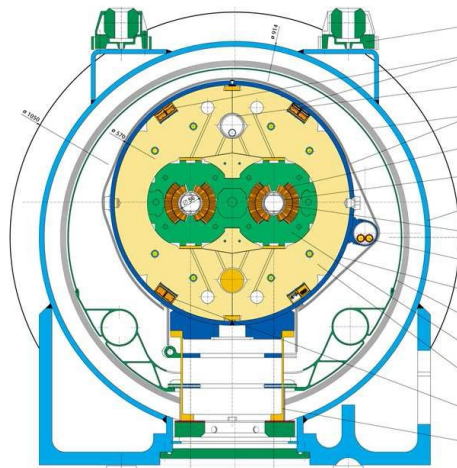


Figure 2.2 – Cross-sectional view of an LHC dipole. The two beam pipes are seen in the center, each surrounded by superconducting coils in orange.

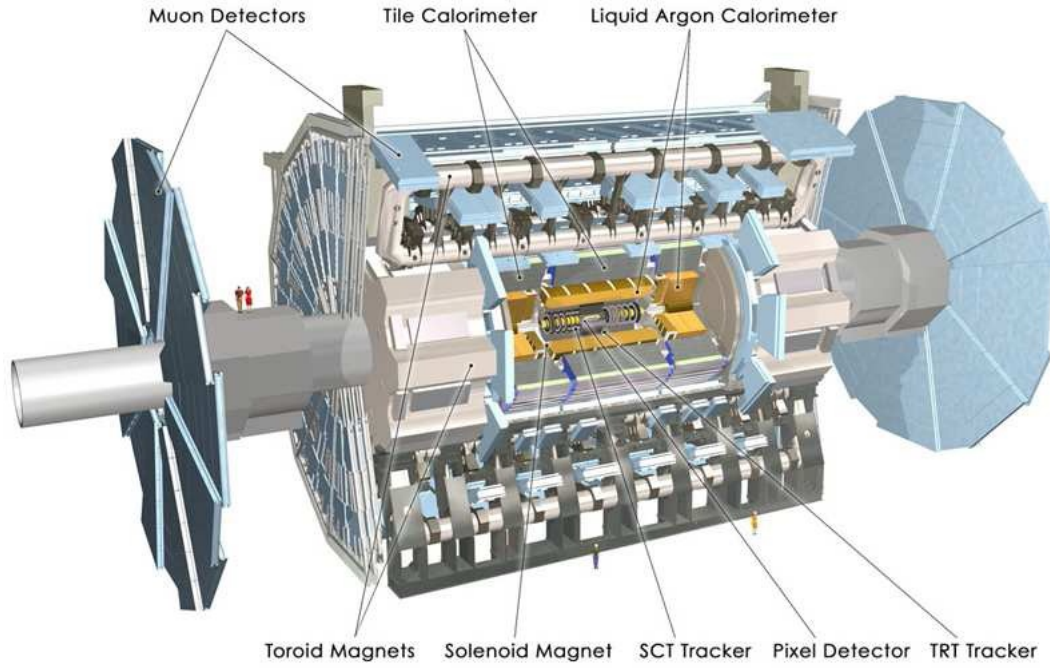


Figure 2.3 – The ATLAS detector

2.2 Introduction to the ATLAS detector

ATLAS is a general-purpose particle detector designed to exploit the full discovery potential of the LHC [19][20][21]. The overall detector has a cylindrical symmetry with a total length of 42 m and a radius of 11 m weighting 7000 tons (see Figure 2.3). The detector is installed 100 m under the ground level at the Interaction Point 1 of the LHC.

The detector is composed by a toroidal magnetic system and five main sub-detectors: the Inner Detector (ID), the electromagnetic liquid Argon calorimeter (LAr), the hadronic calorimeter (Tile) the muon spectrometer (MS) and the forward detectors (Lucid and BCM). The ID tracks the particles trajectory, then the particle energy is

measured by the calorimeters; at the very outer part of ATLAS the MS detects the very penetrating muons. The magnetic system is designed to bend the charged particle trajectory, in order to measure the particle momentum while the forward detectors accomplish the luminosity measurement task.

The standard ATLAS coordinate system considers beam direction as the z-axis, then the x - y plane is transverse to the beam direction; the x-axis pointing to the center of the LHC ring whereas the y-axis points to the surface. The azimuthal angle ϕ is measured around the beam axis, and the polar angle θ is the angle from the beam axis. The pseudorapidity is defined as

$$\eta = -\ln (\tan \theta/2) \quad 2.1$$

and the rapidity as

$$y = \frac{E + p_z}{2(E - p_z)} \quad 2.2$$

The pseudorapidity tends to the rapidity in the limit of a mass-less particle.

2.3 Magnetic system

ATLAS is characterized by two different magnetic field systems required for the identification of charged particles and momentum measurements.

- **Central Solenoid (CS)**, depicted in blue in Figure 2.4, is a super-conducting solenoid providing a magnetic field of 2 T; it is installed around the Inner Detector cavity with a radius of 1.2 m and a length of 5.3 m. It is optimized in order to minimize the amount of material in front of the electromagnetic calorimeter;
- The large super-conducting air-core toroid system is constituted by eight **Barrel Toroids (BT)** and two **End-Cap Toroids (ECT)**, with an open structure to minimize the contribution of multiple scattering to the momentum resolution. It is represented in red in Figure 2.4. Over the range $\eta \leq 1$, magnetic bending is provided by the large barrel toroid, extending over a length of 25 m, with an inner core of 9.4 m and an outer diameter of 20.1 m. For $1.4 < \eta < 2.7$, charged tracks are bent by the two end-cap magnets installed at both ends of the barrel toroid. They have a length of 5 m, an inner core of 1.64 m and an outer diameter of 10.7 m. Over $1 < \eta < 1.4$, usually referred to as the transition region, magnetic detection is provided by a combination of barrel and end-cap. This magnets configuration provides a ~ 4 T field that is mostly orthogonal to the muon trajectories.

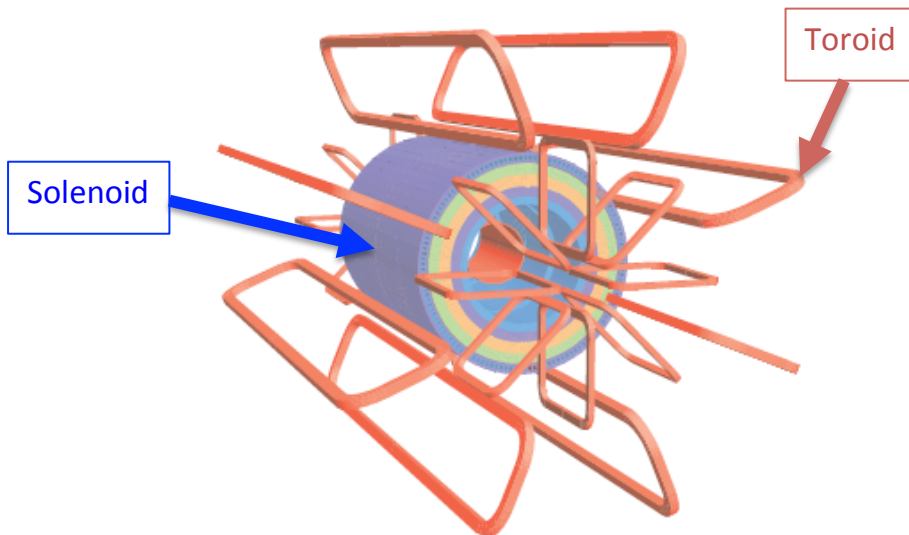


Fig. 2.4 – The Barrel Toroids and the End-Cap Toroids of the magnet system (in red). In blue is there the tile calorimeter steel cylinder

2.4 The Inner Detector

The Inner Detector (ID) is placed in the innermost part of the ATLAS detector. Its structure composed by two silicon revelation systems, the Pixel Detector and the SemiConductor Tracker (SCT), and by a Transition Radiation Tracker (TRT), can be seen in Figures 2.5 and 2.6. The ID is designed to detect the passage of charged tracks and to reconstruct decay vertexes. The high resolution whose tracks are reconstructed by the ID allows, with hadronic calorimeter information, a very precise jet reconstruction. Using additional information from the calorimeter and muon systems, the ID also contributes to electron and muon identification. It is characterized by high momentum resolution and a fine granularity for precise track measurements at a high particle density such as so close to the interaction point.

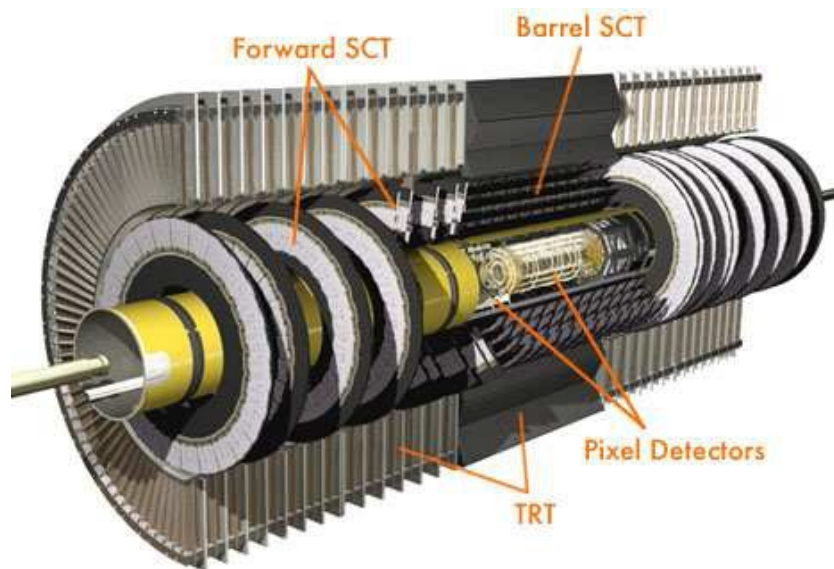


Fig. 2.5 – Cross-sectional view of the Inner Detector.

The technical specifics of the ID systems are labeled in Table 2.1 [22]. The ID system has an acceptance feasible for the analysis purposes of $|\eta| < 2.5$. The detector resolution performance on the impact parameter has been measured using pions; it is $10 \mu m$ in case of pions of $p_T = 5 GeV$ and $35 \mu m$ in case of pions of $p_T = 100 GeV$. The transverse momentum resolution has been found to be equal to 4%, tested with muons having $p_T = 100 GeV$.

Subdetector	Radius [cm]	Element size	Spatial resolution [μm]	Hits/track	Readout channel
Pixel	5 – 12	$50 \mu m \times 400 \mu m$	$10(R - \phi) \times 115(z)$	3	80×106
SCT	30 – 52	$80 \mu m$	$17(R - \phi) \times 580(z)$	8	6×106
TRT	56 – 107	4 mm	130	30	3.5×105

Table - 2.1 – Summary of the main characteristics of the three ATLAS ID systems.

2.4.1 Pixel detector

The Pixel Detector [23] is the nearest detector to the interaction point. It measures the particle impact parameters and the decay vertexes of short living particles such as B hadrons and τ leptons, with a resolution of $15 \mu m$.

A Pixel module is a $16.4 \times 60.8 \text{ mm}^2$ wafer of silicon with 46.080 pixels, $50 \times 400 \mu m^2$ each. The system consists of three barrels at average radii of about 5 cm, 9 cm and 12 cm; to complete the angular coverage five rings are installed on each side, with 11 cm inner radius and 20 cm of outer radius,. Pixel modules in the barrel region are tilted 20° with respect to the cylinder's tangent to counterbalance the effect of the Lorentz

angle. The readout of the pixels, approximately 80.4 million channels, requires the use of advanced techniques. In addition each chip must be radiation-hard to withstand over 300 kGy of ionizing radiation and more than $5 \cdot 10^{14}$ neutrons per cm^2 over ten years of operation of the experiment.

2.4.2 Semi Conductor Tracker

The Semi Conductor Tracker (SCT) system [24] is designed to provide track precision measurements in the intermediate radial range, contributing to the measurement of momentum, impact parameter and vertex position. The SCT barrel consists of 4 concentric layer of radius ranging from 30 to 52 cm from the beam axis. On both sides of the barrel layers, there are 9 SCT disks that cover up to $|\eta| < 2.5$. Each SCT module is made of two strip layers, each of which consists of two 6.4 cm long sensors with a strip pitch of $80 \mu\text{m}$ rotated of 40 mrad with respect one each other. The SCT spatial resolution is $17 \mu\text{m}$ along R- ϕ direction and $580 \mu\text{m}$ in z direction. Tracks can be distinguished if they are separated at least by $\sim 200 \mu\text{m}$.

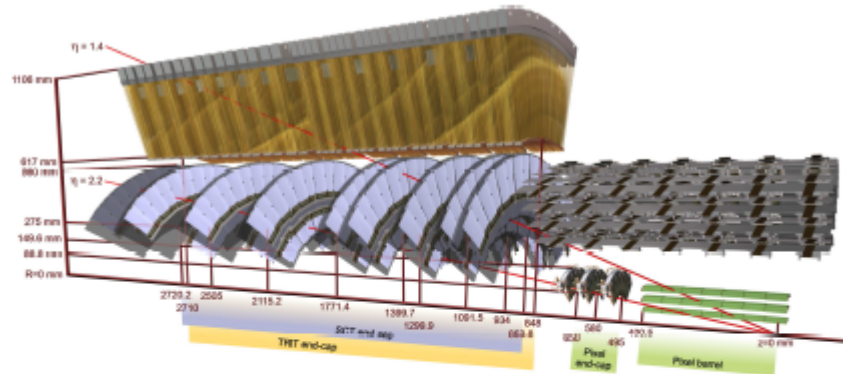


Figure 2.6 – The draw is showing two high η charged tracks traversing the ID elements. It is possible to distinguish Pixel and TRT barrel sensor and the end-cap part of Pixel, TRT and SCT detectors.

2.4.3 Transition Radiation Tracker

The Transition Radiation Tracker (TRT) combines drift tube chamber tracking capabilities with transition radiation detector power of electron/pion discrimination. A single TRT component is composed by a carbon fiber drift tube that cover a 4 mm diameter Kapton straw; this straw contains a 30 μ m diameter gold-plated anode wire. The gap between the straw and the wire is filled by a mixture of gases. The passage of ionizing particle induce a low energy signal on the anodes. At the same time, some particles crossing polypropylene fibers cause transition radiation emission (in the X-ray spectrum) which is absorbed by the Xenon present in the gas mixture. This last process leads to an high energy signal in the TRT electronic that can be distinguished from ionization signal by the voltage intensity. Each straw has a spatial resolution of 130 μ m.

In the barrel region, about 50000 straws are parallel to the beam axis and are 144 cm long, with their wires divided into two halves (approximately at $\eta = 0$) and cover an $\eta < 0.7$ range. In the end-cap region, about 320.000 37cm long straws are arranged radially in wheels, covering $0.7 < \eta < 2.5$ interval. The total number of TRT readout channels is approximately 351.000.

2.5 Calorimetric System

The ATLAS calorimetric system is composed by an electromagnetic Liquid-Argon (LAr) calorimeter covering the pseudo-rapidity region $|\eta| < 3.2$ and an hadronic calorimeter the pseudo-rapidity region $1.7 < |\eta| < 4.9$ (see Figure 2.7). The

calorimetric system structure can be seen in Figure 2.7. Calorimeters must provide good containment for electromagnetic and hadronic showers, and must also limit punch-through into the muon system. Electromagnetic and hadronic calorimeter gives directions and energy information in order to respectively reconstruct electrons and jets that are both fundamentals ingredients for the analyses described in this PhD thesis. In Table 2.2 the specifics for the calorimeter systems are detailed.

Detector component	Energy resolution (σ_E/E)	η coverage
EM calorimeter	$10\% \times \sqrt{1 \text{ GeV}/E} \otimes 0.7\%$	$ \eta < 3.2$ (< 2.5 for the trigger)
Hadronic calorimetry		
Barrel and Endcap	$50\% \times \sqrt{1 \text{ GeV}/E} \otimes 3\%$	$ \eta < 3.2$
Forward	$100\% \times \sqrt{1 \text{ GeV}/E} \otimes 3.1$	$3.1 < \eta < 4.9$

Table 2.2 – Nominal detector performance specifics and coverage for the ATLAS calorimetric system.

2.5.1 Electromagnetic calorimeter

The Electromagnetic calorimeter is divided in a barrel region ($|\eta| < 1.475$) and an end-cap region ($1.375 < |\eta| < 3.2$). The end-cap EM calorimeter is composed by two concentrically wheels. Due to the high radiation level the EM calorimeter makes use liquid argon as active medium and lead as absorber medium.

In the region dedicated to precision measurements (the barrel and the outer end-cap wheel) the EM calorimeter is divided in three longitudinal samplings. The first sampling uses longitudinal strips in the η direction (with a pitch of 4 mm). This sampling is used as a pre-shower detector in order to enhance particle identification and to perform precise position measurement in the η direction.

The second section is segmented in squares of $\Delta\eta \times \Delta\phi = 0.025 \times 0.025$ while the third one has a granularity of $\Delta\eta \times \Delta\phi = 0.05 \times 0.025$. All the calorimeter cells point to the interaction region and the total number of channels is ~ 200.000 . In order to correct for the energy loss in the material before the EM calorimeter, a pre-sampler constituted by an active liquid argon layer between 1 and 0.5 mm is used. In the region between the barrel and the end-cap, the pre-sampler is complemented with a scintillator slab. The energy resolution of the LAr calorimeter has been measured to be less than 1.5%, on analysis performed with photons of energy $E_\gamma = 100 \text{ GeV}$.

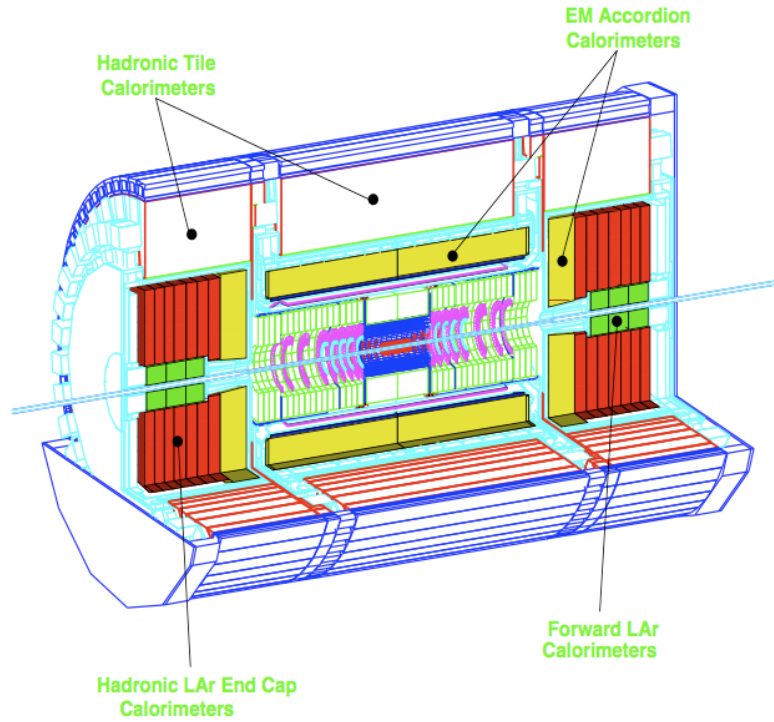


Figure 2.7 – In the picture is shown the position of calorimeter detector. The inner one is the Electromagnetic calorimeter in yellow. In white with red border, is drawn the Hadronic Tile Calorimeter, while the Liquid Argon Calorimeter is shown in red and green.

2.5.2 Hadronic calorimeters

The ATLAS hadronic calorimeter is divided in three different detectors due to the radiation level dependence on the pseudorapidity. In the region $|\eta| < 1.7$ an iron scintillating Tile calorimeter (Tile) is used, for the end-cap and a forward calorimeters Liquid-Argon detectors (HEC) are instead preferred.

The Tile radially extends from 2.28 m to 4.23 m while longitudinally it is made of three layers (1.4, 4.0 and 1.8 interaction lengths). Each calorimetric module is formed of a set of iron tiles partially staggered in the z direction. The void space between the iron tiles is filled with scintillating tiles. The HEC is composed of two wheels of 2.03 m outer radius. The wheel near the interaction point is composed by 25 mm copper plates, while the wheel farther from the interaction point is composed by 50 mm copper plate as a cost saving solution. In both wheels the 8.5 mm gap between the various plates is filled with liquid argon and divided in four gaps by three parallel electrodes. A Liquid Argon Forward Calorimeter (FCAL) covers the pseudo-rapidity region $3.1 < |\eta| < 4.9$ and is located at a distance of 4.7 m from the interaction point; the FCAL is composed by three sections, a first copper section and two tungsten ones.

The ATLAS hadronic calorimeter system has been designed with a thickness of about 10 interaction lengths. The large coverage in pseudo-rapidity allows good E_T^{miss} measurements. The Tile hadronic calorimeter performance on the measurements of the jet energy has been evaluated to be $\frac{\Delta E}{E} = \left(65/\sqrt{E} \oplus 2 \oplus 5/E\right)\%$.

2.6 Muon spectrometer

The Muon Spectrometer (MS) is the outermost ATLAS sub-detector measuring muons momentum in a pseudo-rapidity region of $1 \leq |\eta| \leq 2.7$. The MS structure can be seen in Figure 2.8 where the different systems of which it is composed are placed in evidence: two trigger chambers, the Resistive Plate Chambers (RPC) and the Thin Gap Chambers (TGC), and two high-precision tracking chambers, the Monitor Drift Tubes (MDT) and the Cathode Strip Chambers (CSC). The trigger chambers are fast detectors that perform coarse measurements of muon momentum, while precision chambers measure momentum with a better resolution but with a longer signal build-up. A sophisticated optical alignment system, that can be seen in Figure 2.9, have been designed to meet the stringent requirements on the mechanical accuracy and the survey of the muon chambers. The MS triggers for muon objects that are, expertly if they possess high p_T , indicative of the signature for many events category of physical interest; it is also deputed to the muon reconstruction and momentum evaluation.

The MS performance on the transverse momentum measurement has been evaluated using muons of both $p_t = 10 \text{ GeV}$ and $p_t = 100 \text{ GeV}$ founding a resolution value of 3% and 12% respectively.

The RPC is a gaseous parallel electrode-plate detector organized in strip elements that combines an adequate spatial resolution of 1 cm with an excellent time resolution of 1 ns. The number of strips (average strip pitch is 3 cm) per chamber is variable: 32, 24 or 16 in η and from 64 to 160 in ϕ . When a particle goes through an RPC chamber, the primary ionization electrons are multiplied into avalanches by a high electric field of typically 4.9 kV/mm. The signal is read out via a capacitive coupling of strips on both

sides of the chamber. The TGC are very thin multi-wire proportional chambers which peculiarity is to have cathode-anode spacing smaller than the anode-anode (wire-wire) spacing. This characteristic justify the denomination *Thin Gap* and allows a very short drift time and an excellent response in time, less than 20 ns; this meets the requirement for the identification of bunch crossings at 40 MHz. The TGC are filled with a highly quenching gas mixture of 55% of CO_2 and 45% of n-pentane (C_5H_{12}).

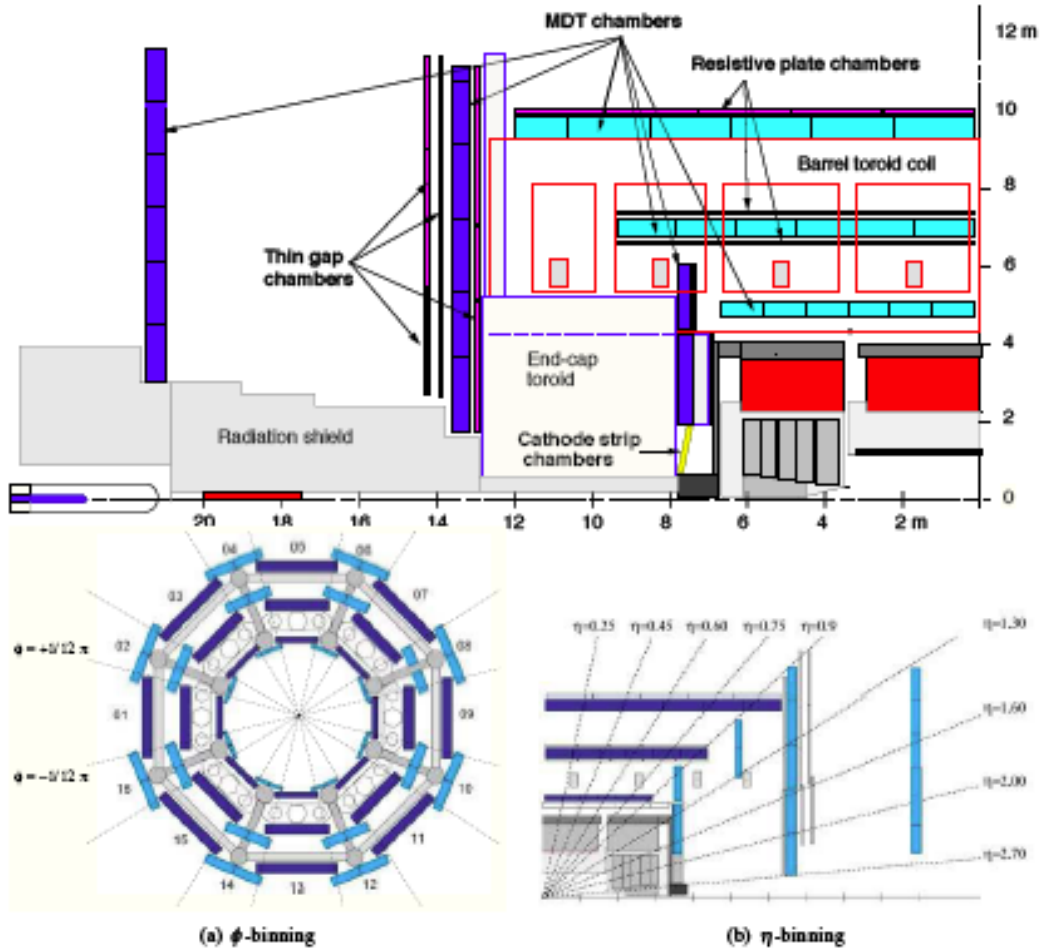


Figure 2.8 – Muon Spectrometer layout (top) and his binning in η (down-left) and ϕ (down-right) projections.

The MDT system is composed by drift chambers consisting of two multi-layer drift tubes (each multi-layer has three or four layer of tubes). The difference between MDT and traditional drift chambers is that each drift cell is enclosed in an aluminum tube, which provide mechanical stability to the chambers. A single MDT tube resolution is of $80\ \mu\text{m}$, while the all MDT system reach a resolution of $35\ \mu\text{m}$ along the ϕ direction.

The CSC chambers are proportional multi-wire chamber with segmented read out cathode. The drift cells are symmetric, i.e. the distance between the anode wires and the cathode is equal to the intra wire distance ($2.54\ \text{mm}$). The cathodes are segmented in $1\ \text{mm}$ strips, orthogonal to the anode wires, allowing to the measure the crossing point of incoming muon with resolution of $40\ \mu\text{m}$ in the ϕ direction. In the η direction the cathode segmentation is coarser leading to a resolution of $5\ \text{mm}$.

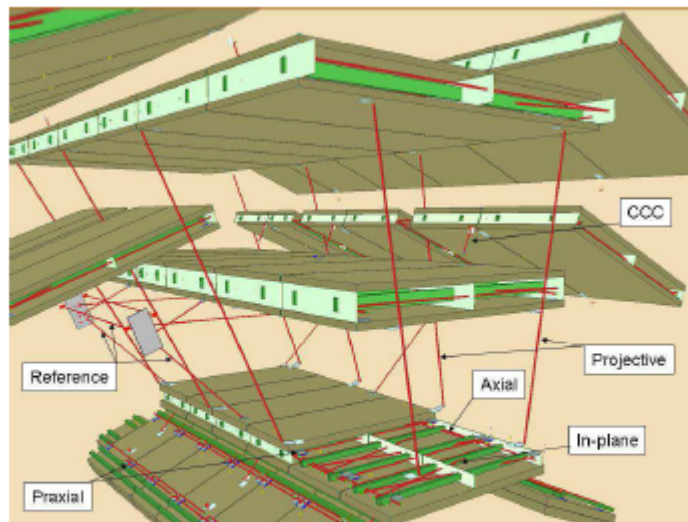


Fig. 2.9 – 3D representation of the Muon Spectrometer and his optical-alignment system.

2.7 Luminosity measurement

The luminosity measurement, necessary for the cross section evaluation, has been performed by the LUCID, BCM, FCAL and TILECAL detectors: all the details about its determination can be found in [82] [83], here only a brief description of the LUCID and BCM detectors is given (FCAL and TILECAL have been already faced) just to better understand all points of the analysis. The luminosity has been evaluated with a precision of 2.8% (3.7%) for the data acquired in 2011 (2012).

2.7.3 Lucid

LUCID (Luminosity Cherenkov Integrating Detector) is a Cherenkov detector specifically designed to measure the luminosity in ATLAS. It is composed by two identical parts located around the beam pipe at 17 m from the interaction point covering a pseudo-rapidity range $5.6 < |\eta| < 6.0$. Each part is composed by 16 aluminium tubes filled with the C₄F₁₀ gas. A charged particle travelling the gas inside the detector emits some Cherenkov photons that after various reflection by the tube walls reach a photomultiplier (PMT) situated at the back end of each tube. Additional Cherenkov photons are produced in the quartz window of the PMT for a total of about 100 photoelectrons per incident charged particle. Each tube is considered hit if it receives a charge over a preset threshold equivalent on average to 15 photoelectrons. Through different algorithms it is possible, from the number of hit tubes, to evaluate the average number of interactions per bunch crossing and then the luminosity. LUCID

gives a relative luminosity measurement for each bunch crossing monitoring its behavior in order to compare its results with the BCM detector and to indicate eventual problem of LHC.

2.7.2 BCM

The BCM detector consists of two stations located symmetrically around the interaction point covering a pseudo-rapidity region of about 4.2. Each station is composed by four modules in which there are two diamond sensors read out in parallel. The electronic is very fast in reading to provide a measurement of bunch-by-bunch luminosities in ATLAS. It was indeed originally designed to monitor background levels and issue beam-abort requests when beam losses start to risk damaging the Inner Detector.

2.8 Minimum Bias Trigger Scintillator

The Minimum Bias Trigger Scintillator (MBTS) is especially used to trigger *minimum bias* events, which have the only requirement to be generated by a *pp* collision, in low luminosity runs. The MBTS counters consist of one plane with 16 scintillator segments on each side of the interaction point, mounted in front of the end-cap calorimeter and connected to different photomultiplier tubes. There are 2 segments to

measure the η direction (inner and outer) and 8 segments for the θ direction. The MBTS detector covers a pseudo-rapidity range of $1.9 < |\eta| < 3.8$ and has an overall acceptance on inelastic interaction of about 80%.

2.9 Trigger System

The trigger system task is to reduce the initial bunch-crossing rate from ~ 40 MHz, to about 200 Hz, necessary for the permanent storage of the data; each data stream has a different trigger chain based on specific sub-detectors and algorithms. In this analysis we are interested in the *Muons* and the *Egamma* data streams. Triggers are used in parallel, meaning that an event is accepted if at least one trigger chain is satisfied; this imply the necessity of further selection cuts during the offline analysis to reject events selected by triggers not specific for particular analysis on which we are interested for.

This goal is performed by a large number of trigger decision tools organized in three levels called Level1 (L1), Level2 (L2) and Event Filter (EF) and containing algorithm of increasingly complexity and elaboration time (see Fig. 2.10). Here I will give a brief description of the three trigger levels implemented for the muon pair analysis.

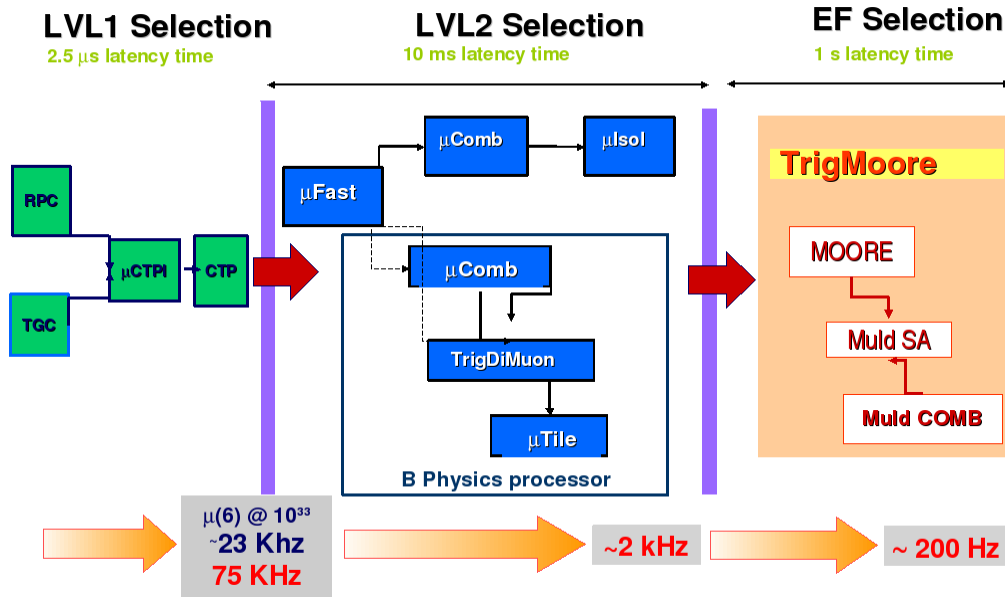


Fig. 2.10 – ATLAS trigger chain with the different algorithm used in each level.

Level1 trigger algorithms are studied to discard the major part of non interesting events with the minimum CPU time possible. After Level1 selection, the data acquisition rate is decreased to $\sim 75 \text{ kHz}$. Level2 trigger is a software-based system, with selection algorithms running on a farm of commodity PCs. The selection is largely based on regions-of-interest (RoI) identified at L1 and uses fine-grained data from the detector for a local analysis of the L1 candidate. A seed is constructed for each trigger accepted by L1 that consists of a p_T threshold and an $\eta - \phi$ position. The L2 algorithms use this seed to construct a RoI window around the seed position. Event Filter triggers are the last step before data recording and they have to further refine events passed the previous selection.

Chapter 3

Data samples and Monte Carlo models for Signal and Background

3.1 Collected Data Samples

The two described in this thesis analyses has been performed on the data collected by the ATLAS detector in 2011 and 2012 respectively at $\sqrt{s} = 7 \text{ TeV}$ and $\sqrt{s} = 8 \text{ TeV}$.

Events accepted for the 2011 analysis must satisfy stable beam conditions and all detector systems must had worked properly during the acquisition period; the

corresponding integrated luminosity is 4.66 fb^{-1} . Luminosity calibrations has been determined from Van der Meer scans performed during the data taking, with a relative uncertainty of 3.4% [32]. The data used in this analysis belongs to different periods corresponding to different LHC conditions. Data have been furthermore divided in an electron (*Egamma*) and a muon (*Muons*) stream depending on the selected online triggers; these streams are analyzed separately in the $e + jets$ and $\mu + jets$ channels. The trigger selection used change depending on the data period to account for the continuous increasing of the instantaneous luminosity that leads to a corresponding increasing of the pile-up events. The trigger used are:

- ❖ Single electron triggers (EF_e20_medium, EF_e22_medium, EF_e22vh_medium1) requiring a reconstructed “medium” electron of p_T greater than 20 or 22 GeV. Medium electron are defined by the following requirements: $|\eta| < 2.47$, a certain ratio between energy deposit in EM and hadronic calorimeter, at least one hit in the detector pixel, and at least 7 hits in the SCT pixels; a minimum impact parameter of 5 mm is also required. The letters vh were added for those triggers seeded by L1 items with η -dependent thresholds and a hadronic leakage requirement[73] [74];
- ❖ Single muon triggers (EF_mu18, EF_mu18_medium) requiring a reconstructed generic or “medium” muon of p_T greater than 18 GeV. The definition of medium muons requires acceptance cuts and a certain number of hits in MS chambers and ID pixels depending on the type of reconstruction used for the muon[74].

Period	Integrated Luminosity (pb^{-1})	Electron Trigger	Muon Trigger
B	17	EF_e20_medium	EF_mu18
D	179		
E	50		
F	152		
G	560		
H	278		
I	399	EF_e22_medium	EF_mu18_medium
J	232.9		
K	660.2		
L	1568	EF_e22vh_medium1	
M	1121		

Table 3.1 – Luminosity collected and event filter triggers used in data taking period for the complete 2011 data set.

The periods and the triggers used in the analysis, together with their integrated luminosity, are reported in Table 3.1

The analysis performed on 2012 data has used the full set of events collected by the ATLAS detector in pp collisions at an increased center of mass energy $\sqrt{s} = 8TeV$ and corresponding to an integrated luminosity of $\int \mathcal{L}dt = 20.344fb^{-1}$. In Figure 3.1 the luminosity collected in 2012, together with the 2011 one, is shown as function of time; here it is possible to notice the luminosity increase between the two data taking. Electron and muon EF triggers are the same for all periods, respectively.

- EF_e60_medium1 OR EF_e24vhi_medium1 asking for an isolated “medium” electron with $p_T > 60 \text{ GeV}$ or with $p_T > 24 \text{ GeV}$ in the electron channel;
- EF_mu24i_tight OR EF_mu36_tight asking for an isolated “tight” muon with $p_T > 36 \text{ GeV}$ or with $p_T > 24 \text{ GeV}$ in the muon channel. Tight muon requirements are similar to the medium requirements but more stringent; additional cuts on momentum and p_T are done depending on the reconstruction used for the muon.

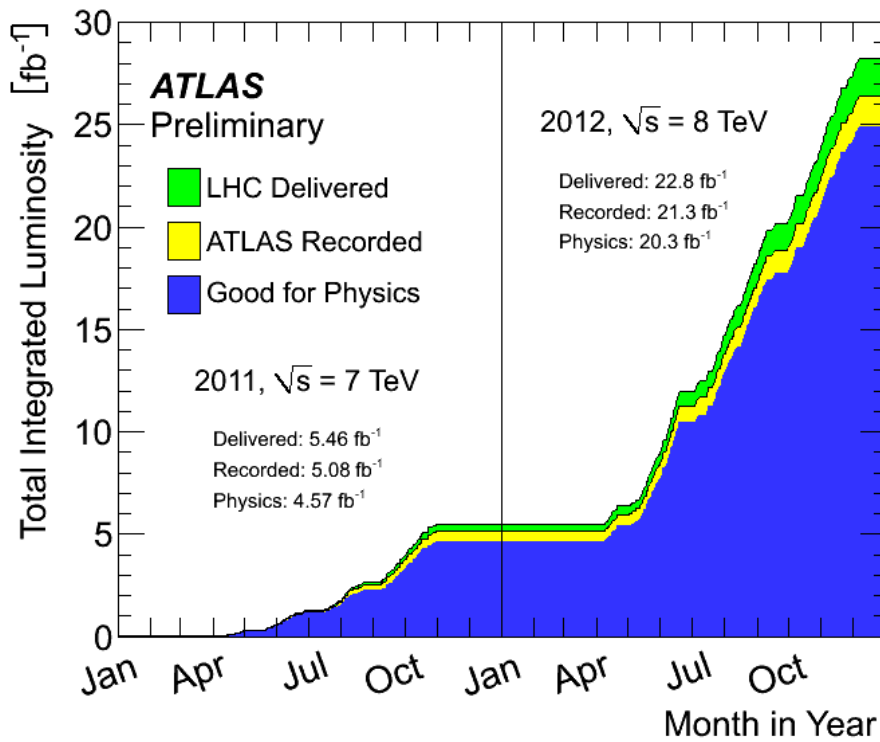


Figure 3.1 – Luminosity as a function of time collected during the 2011 and 2012 data takings [96].

3.2 Monte Carlo Simulation

The physical processes and the interactions of the final state particles with the detector are simulated by Monte Carlo programs; the produced MC samples are generated according to both theoretical predictions and phenomenological models. These events are used to test our comprehension of the physics processes and the experimental apparatus behavior by a comparison with the experimental data. The Monte Carlo simulation is usually divided in two steps.

The event generation is performed by theoretical calculations of the elementary processes from the pp interaction to the final state stable particles, following all intermediate steps. The first generation step is the calculation, at a fixed perturbative order, of the hard process (parton-parton interaction) matrix element, followed by the QCD cascade generation, called parton shower, (a simplified diagram of the gluon emission process in a generic process is exemplified in Figure 3.2). The parton shower is a *space-like* process for the initial state partons but a *time-like* process when applied to the final-state partons. In the initial state the QCD radiation emission progressively increases the virtuality of the initial state partons, allowing them to access to the hard scale needed to describe perturbatively the hard scattering process. After the scattering, the time-like parton shower process leads to the emission of gluons from the produced particles. Once the final partons are generated, phenomenological hadronization models are used in order to produce the stable particles in the final state. These final particles are

passed through the detector simulator. A pictorial representation of the event generation chain just described is shown in Figure 3.3.

The detector simulation is performed by the GEANT4[33] simulator that provides a model for the particle interaction through matter; for this reason a detailed description of the ATLAS detector geometry and of the trigger system is necessary. Due to computer elaboration time necessities, some Monte Carlo sample has not been processed using the full detector simulation but with an approximate and faster simulator, namely *AtlasFastII* (AFII).

MC events after the detector simulator processing can be reconstructed and analyzed with the very same code used for the collision data.

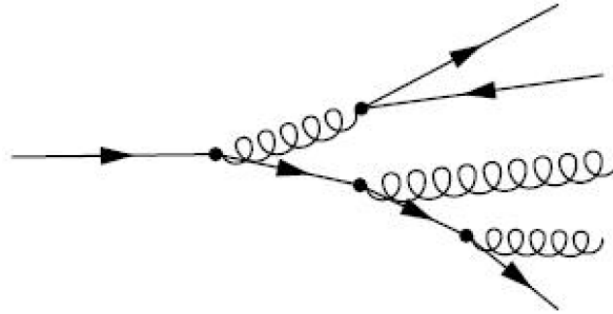


Figure 3.2 – Example of a parton showering diagram.

3.3 $t\bar{t}$ Monte Carlo Signal

The simulation of the $t\bar{t}$ production for 2011 data taking has been performed using two different Monte Carlo generators, ALPGEN[34] and MC@NLO[35]; both of them set the top quark mass to $m_t = 172.5\text{GeV}$.

The ALPGEN sample contains only the simulation of $t\bar{t}$ pairs where one top decays hadronically and the other one leptonically ($\ell + jets$ channel); the sample is divided in different subsamples depending on the number of additional final state partons generated. The ALPGEN generator accounts only for leading order Feynman diagrams using the CTEQ6L1[36] LO set of PDFs. The parton shower is instead simulated by Herwig[37] and JIMMY[38].

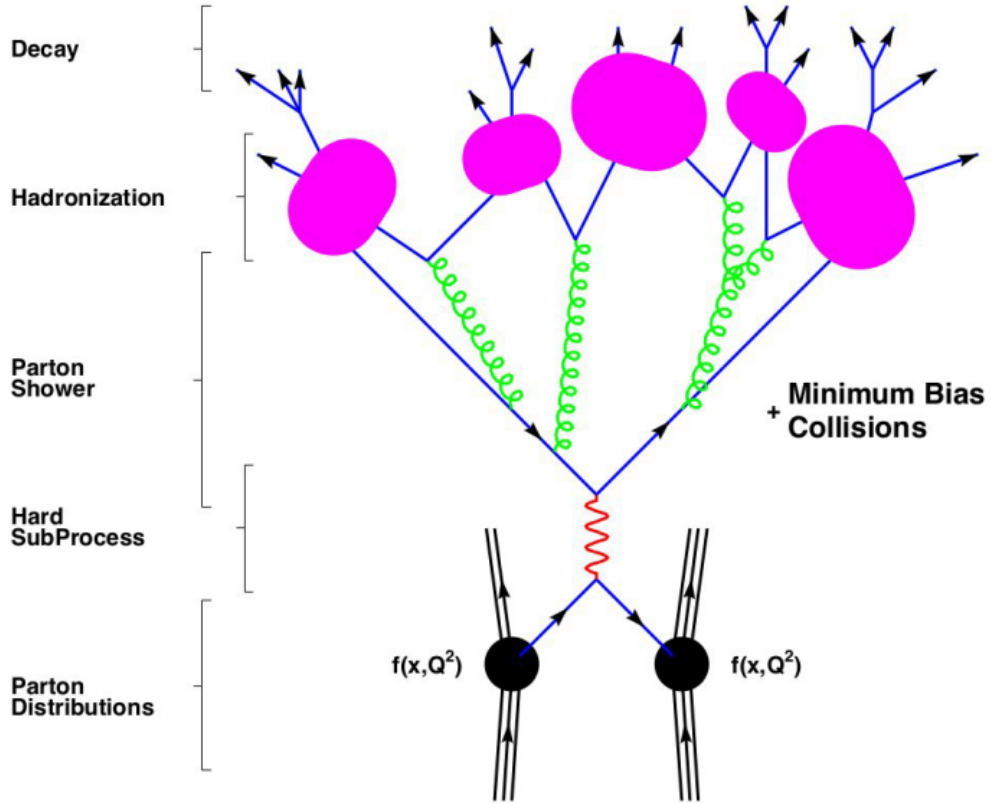


Figure 3.3 – Pictorial representation of a Monte Carlo event generation chain

MC@NLO $t\bar{t}$ events are generated taking into account NLO diagrams and then are rescaled to the approximated NNLO calculation performed by HATHOR[39] with a proper k -factor. The CT10[40] PDF set at the NLO is used. The MC@NLO sample contains both *dilepton* (where both top quarks decay leptonically) and $\ell + jets$ events, which contribution will be separated during the analysis.

With respect to MC@NLO, ALPGEN provides a better agreement with collected data in high jet multiplicity distributions as shown in Figure 3.4; for that reason ALPGEN has been chosen as default signal generator in the 2011 data analysis, while MC@NLO has been used for systematic checks.

For the 2012 data taking simulation, two different Monte Carlo samples have been taken into account for the $t\bar{t}$ signal simulation, both of them including $\ell + jets$ and *dilepton* events. The baseline sample has been generated by the NLO framework Powheg[41] using the CTEQ6L1 PDF set, followed by the Pythia[42] parton shower simulation. The second MC sample, used for systematic checks, has been produced with the MC@NLO generator, the C10 PDF set and HERWIG+JIMMY for the parton showering simulation.

3.4 Monte Carlo and Data Driven Background sources

Background sources affecting the $t\bar{t}$ pair production channel are constituted by those physical processes leading to a final signature similar to the $t\bar{t}$ one. This could be due both to the similarity of the decay products themselves or to the not negligible

probability of objects misidentification by the detector. These backgrounds can be simulated as using standard Monte Carlo models as via specifics *data driven* techniques. *Data driven* techniques are those that evaluates the background contributions directly from acquired data, taking into account only appropriate control regions. The processes giving larger background contribution to the signal $t\bar{t}$ pair in the $\ell + jets$ channel are single top production, multijets events from QCD processes, $Z + jets$, $W + jets$ and diboson events. Single top, $Z + jets$ and diboson background are evaluated only by MC simulation while $W + jets$ Monte Carlo background distribution requires an overall normalization plus an heavy flavor rescaling, both calculated via data driven methods; the QCD background is entirely evaluated using a *data driven* technique. In Table 3.2 the production cross section for the signal and the background samples considered in both the analyses respectively performed at a center of mass energy $\sqrt{s} = 7 TeV$ and $\sqrt{s} = 8 TeV$ are reported.

SINGLE TOP

The background from electroweak single top quark production is about a factor of two smaller than the $t\bar{t}$ cross section (see Chapter 1) and due to the lower number of jets with respect to the $t\bar{t}$ production, it contributes predominantly in low multiplicity events. This background is simulated using MC@NLO with the CT10 PDF set for what concerning the s- and Wt-channels while AcerMC[43] and Pythia, with the addition of the MRST2007/LOMOD set of PDF, are used respectively for the event generation and for the parton shower in the t-channel case.

Z + jets

Z + jets events can be mismatched for $t\bar{t}$ processes in both the electron and muon Z boson decays ($Z \rightarrow e^+e^-$ and $Z \rightarrow \mu^+\mu^-$), where one lepton is not detected giving the needed fake E_T^{miss} contribution, and in the tau decay case ($Z \rightarrow \tau^+\tau^-$), where one τ lepton decay leptonically and the other hadronically. This source of background is estimated with the ALPGEN+Herwig/Jimmy MC generators.

DIBOSON

A small background contribution is given by diboson events $pp \rightarrow WW, WZ, ZZ$ which decay products can have the same final configuration as in $t\bar{t}$ events.

QCD

This totally data driven background will be described in Paragraph 3.4.1.

W+JETS

$W + jets$ events constitutes the main background source for both the 2011 and 2012 data taking analyses presented in this thesis because of the high cross section and the signature very close to the $t\bar{t}$ one, especially in the high jet multiplicity case. This source of background will be described in Paragraph 3.4.2.

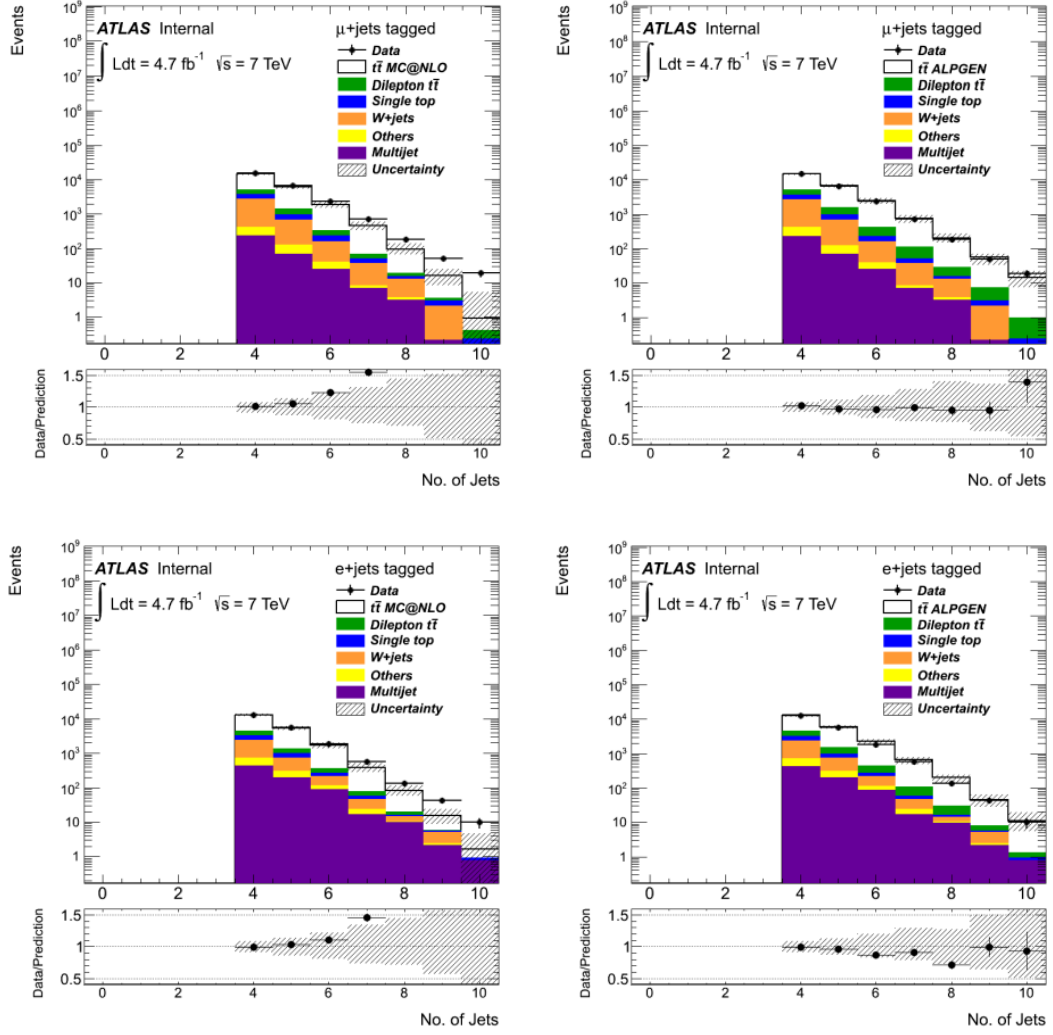


Figure 3.4 – Number of jets spectra in the electron, on top, and muon, on bottom, channels after the one-tag selection (as described in Chapter 5) using the MC@NLO generator (on the left) and the Alpgen generator (on the right) to model the signal.

MC sample	7 TeV Cross Section (pb)	8 TeV Cross Section (pb)
Powhed+Pythia $t\bar{t}$ signal	—	137.32
MC@NLO $t\bar{t}$ signal	96,31	137,31
Alpgen $t\bar{t}$ signal	96,27	—
Single top	~ 55	~ 72
Z+jets	~ 7	~ 33
Diboson	~ 1070	~ 1148
W+jets	$\sim 30 \cdot 10^3$	$\sim 37 \cdot 10^3$

Table 3.2 – Cross Section, corrected for the k -factor, used in Monte Carlo production both for signal and background source. In case of $t\bar{t}$ samples the not-fully-hadronic cross section is shown.

3.4.1 QCD or Fake Leptons background

An important background source is due to QCD multi-jets events in correspondence to a lepton misidentification by detector that deceives the single lepton triggers. The objects much commonly identified as “fake leptons” are long living mesons (i.e. π^\pm or K^\pm), photons and hadronic jets. The misidentification rate is very small but due to the huge multi-jets cross section the contribution is not negligible. This source of background is usually called QCD or fake-leptons background. The QCD background is highly detector dependent, hence the better way for its estimation is via data driven methods; the one used in both the 2011 and 2012 analyses is the Matrix Method (MM). It has already been used by the CDF and D0 experiments at Tevatron[44] and is based on

the determination of the efficiency between signal-like and a fake-like events selected with different lepton requirements.

The first step of the Matrix Method consists in the selection of two different event samples, called “tight” and “loose”, that differ only in the lepton cut definition. The tight selection is exactly the one applied in standard analysis while the loose events are selected with a looser requirement in the leptonic cut; in the specific case of the analyses presented in this thesis, the lepton isolation cut has been removed in the loose selection.

The number of events surviving to the tight and loose (N^{tight} and N^{loose}) selection can be expressed as the sum of the correspondent number of real (true) signal events ($N_{real}^{tight(loose)}$) plus the number of events from the lepton misreconstruction ($N_{fake}^{tight(loose)}$):

$$N^{tight} = N_{real}^{tight} + N_{fake}^{tight} \quad 3.1$$

$$N^{loose} = N_{real}^{loose} + N_{fake}^{loose}$$

The equation system 3.1 can be more conveniently rewritten as

$$N^{tight} = \varepsilon_{real} N_{real}^{loose} + \varepsilon_{fake} N_{fake}^{loose} \quad 3.2$$

defining the real and fake efficiencies as

$$\varepsilon_{real} = \frac{N_{real}^{tight}}{N_{real}^{loose}}$$

3.3

$$\varepsilon_{fake} = \frac{N_{fake}^{tight}}{N_{fake}^{loose}}$$

Now it is easy to retrieve the number of fake-lepton background defined as

$$N_{fake}^{tight} = \frac{\varepsilon_{fake}}{\varepsilon_{real} - \varepsilon_{fake}} (\varepsilon_{real} N^{loose} - N^{tight}) \quad 3.4$$

The ε_{real} is estimated via a tag-and-probe technique from a sample of $Z \rightarrow e^+e^-$ events that do not contains fake-lepton contamination. The same selection applied in the standard analysis case is used, except for jet dependent requirements and for the overlap removal, which is also applied in the loose case. Events with at least a tagged lepton (request to be tight) are the baseline sample to be used in order to determine the efficiency that is evaluated searching for a second, probe, lepton.

The fake efficiency ε_{fake} is evaluated on a $t\bar{t}$ signal sample with the requirement of at least one jet with $p_T > 25 \text{ GeV}$, exactly one loose lepton, a missing transverse energy of $E_T^{miss} < 20 \text{ GeV}$ and a minimum distance between the highest p_T jet and the lepton $\Delta R_{leading-jet,lepton} \geq 0.7$. The efficiency is evaluated as the ratio of events in which the selected loose lepton also pass the tight requirements, divided by the total number of loose events. Events with a different E_T^{miss} cut are used in the determination of the systematic uncertainty of the matrix method.

The fake-lepton efficiencies are $\eta - p_T$ dependent, so a 2-dimensional efficiency map is produced to retrieve the weight to be applied to real data events, obtaining the data-driven QCD background estimation. The weight is defined as

$$\varpi_i = \frac{\varepsilon_{fake}}{\varepsilon_{real} - \varepsilon_{fake}} (\varepsilon_{real} t_i - (1 - t_i)) \quad 3.5$$

where t_i is equal to 1 if the event i is loose and to 0 if it is tight. This weigh should be applied to data sample in order to estimate the background contribution from QCD multi-jet events.

3.4.2 $W + jets$ background

The events estimated using theoretical $W + jets$ cross section entails large uncertainties, so two different data-driven techniques are used in order to reduce Monte Carlo uncertainties: a global normalization obtained through the *charge asymmetry* method and a correction for the different Heavy Flavour Fractions (HFF) of jet components. The *ALPGEN+Pythia* Monte Carlo has been used to simulate the $W + jets$ production.

CHARGE ASYMMETRY METHOD

A global normalization scale factor is estimated using the *charge asymmetry data-driven* technique based on the non-symmetric distribution of W^+ and

W^- in the $W + jets$ events from pp collisions. The W^+ bosons can be produced from parton level processes such as $u\bar{d} \rightarrow W^+$ or $c\bar{s} \rightarrow W^+$ while W^- bosons production can be $d\bar{u} \rightarrow W^-$ or $s\bar{c} \rightarrow W^-$. The proton PDFs, from which the cross section depends, is different for the various quarks, i.e. the interaction probability of u-quarks, coming from its PDF, is greater than the d-quarks one; so the cross section of $u\bar{d}$ processes $(\sim u(x_1) \cdot \bar{d}(x_2))$ results wider than the $d\bar{u}$ one $(\sim \bar{u}(x_1) \cdot d(x_2))$. Hence there is an asymmetric production of different sign W bosons that leads to an asymmetric lepton charge distribution. The $t\bar{t}$ process and the other backgrounds considered in these analyses are charge symmetric, except for the single top and the diboson contribution. Once these backgrounds are subtracted from data, the only remaining charge asymmetry source is from $W + jets$. This allows making the following assumption

$$N_{W^+} - N_{W^-} \approx N_{D^+} - N_{D^-} \quad 3.6$$

where $N_{D^{+(-)}}$ is the number of data events with a positive (negative) lepton and $N_{W^{+(-)}}$ is the number of positive (negative) W from simulated $W + jets$ events. The total number of $W + jets$ events can be evaluated from equation

$$\begin{aligned} N_W &= N_{W^+} + N_{W^-} = \frac{N_{W^+}^{MC} + N_{W^-}^{MC}}{N_{W^+}^{MC} - N_{W^-}^{MC}} \cdot (N_{D^+} - N_{D^-}) \\ &= \frac{R + 1}{R - 1} \cdot (N_{D^+} - N_{D^-}) \end{aligned} \quad 3.7$$

where the cross section ratio $R = \sigma(pp \rightarrow W^+X)/\sigma(pp \rightarrow W^-X)$ is relatively well understood and can be estimated from Monte Carlo simulation [45]. This procedure allows a reduction of the total $W + jets$ cross section uncertainty with respect to the only MC estimation.

HEAVY FLAVOUR FRACTION

The $W + jets$ background is composed by heavy flavour components ($Wb\bar{b}$, $Wc\bar{c}$, Wc) and a light flavour one ($W\ell$). The determination of heavy flavour fraction (HFF) sources suffers of large theoretical uncertainties, as well as for the overall normalization. These uncertainties have a strong impact on the final samples selected after b-bagging requirements, so it is important to decrease them using a *data-driven* estimation. The number of b-tagged events $N_{i-jet}^{W,tag}$ can be expressed for each jet multiplicity sample i as a function of the number of events before b-tagging $N_{i-jet}^{W,pretag}$ (pre-tagging). The b-tagging probabilities $P_x(x = b\bar{b}, c\bar{c}, c, \ell)$ for each flavour component can be estimated by Monte Carlo studies. The flavour fraction of pre-tagged events $F_{b\bar{b}}$, $F_{c\bar{c}}$, F_c and F_ℓ , are the quantities needed to be measured:

$$N_{i-jet}^{W,tag} = N_{i-jet}^{W,pretag} (P_{b\bar{b}}F_{b\bar{b}} + P_{c\bar{c}}F_{c\bar{c}} + P_cF_c + P_\ell F_\ell) \quad 3.8$$

The HFF sum must be equal to 1

$$F_{b\bar{b}} + F_{c\bar{c}} + F_c + F_\ell = 1 \quad 3.9$$

and the number of $W + jets$ events in each jet multiplicity sample is equal to the number of events in data after the non-W component subtraction:

$$N_{i-jet}^W = N_{i-jet}^{data} - N_{i-jet}^{MC\ non-W} - N_{i-jet}^{QCD} \quad 3.10$$

The method first estimates HFF from the 2-jet multiplicity sample, the one with higher statistic and lower uncertainties. Equation 3.8 becomes

$$N_{2-jet}^{W,tag} = N_{2-jet}^{W,pretag} (P_{b\bar{b},2} F_{b\bar{b},2} + P_{c\bar{c},2} k_{c\bar{c} \rightarrow b\bar{b}} F_{b\bar{b},2} + P_{c,2} F_{c,2} + P_{\ell,2} F_{\ell,2}) \quad 3.11$$

by substituting

$$F_{c\bar{c}} = k_{c\bar{c} \rightarrow b\bar{b}} F_{b\bar{b}} \quad 3.12$$

where $k_{c\bar{c} \rightarrow b\bar{b}}$ is the ratio between $Wc\bar{c}$ and $Wb\bar{b}$ from Monte Carlo. Processing separately the W^+ and W^- events and imposing that the number of b-tagged W-events should be the same in data and MC, it is possible to evaluate the flavour fractions. These values are applied to MC events and the procedure is iteratively repeated until no significance variations are observed. The same procedure is applied to higher jet multiplicity samples starting from HFFs found from the 2-jet multiplicity bin.

Chapter 4

Object Reconstruction

The foundation of any ATLAS physics analysis is constituted by reconstructed and identified objects that represent the observed characteristics of the particles produced by the pp interactions and travelling through the detector volume. A description of the reconstructed object used in the $\ell + jets$ analyses presented is the argument of this chapter and comprehends jets of different radius, muons, electrons and missing transverse energy from neutrinos.

4.1 Jets

The basic structures of the jet reconstruction process in ATLAS are locally-calibrated, three-dimensional topological clusters (*topo-clusters*), built from

calorimeter's cells [46]. Topo-clustering reconstruction starts with the identification of seed cells that should have energy significance at least 4σ above noise level, where the noise is defined as the sum in quadrature of electronic and pile-up signals. Neighbor cells with energy significance higher than 2σ with respect to the noise, are then iteratively added to form the clusters. An extra ring of direct neighbor cells is finally added to the clusters. After *topo-clusters* reconstruction, a splitting algorithm divides clusters in energy-categories using a local energy maxima criterion. Individual clusters are calibrated using local properties such as energy density, calorimeter depth and isolation with respect to nearby clusters. This local cluster weighting (LCW) calibration is derived from dedicated single pion Monte Carlo simulations and allows to classify clusters as electromagnetic or hadronic.

In these analyses jets are reconstructed using the anti- k_t algorithm [47], a method based on a sequential cluster recombination algorithm that follows an iterative procedure; the intermediate reconstruction objects are called "*pseudo-jets*". The description of the method follows:

1. Define a distance d_{ij} for each pair of objects (clusters or *pseudo-jets*) i, j and the $beam(B)$

$$d_{ij} = \min(p_{T,i}^{2k}, p_{T,j}^{2k}) \frac{\Delta R_{ij}^2}{R^2} \quad 4.1$$

$$d_{iB} = p_{T,i}^{-2}$$

with

$$\Delta R_{ij}^2 = (\eta_i - \eta_j)^2 + (\varphi_i - \varphi_j)^2 \quad 4.2$$

The parameter k is fixed equal to -1 ; $p_{T,i}$, η_i and φ_i are respectively the transverse momentum, the pseudorapidity and the azimuthal angle of the i object; R is an input parameter that limits the cone radius of the jet;

2. Find the minimum distance among d_{ij} and d_{iB} ;
3. If the minimum value is d_{ij} then combine i and j into a single *pseudo-jet*.

Otherwise consider i as a final state and do not consider it in further iterations. Repeat from step 1.

In the following, the jets reconstructed by the anti- k_t algorithm (parameter $k = -1$) will be used; this choice favours the clusterization around hard particles rather than soft ones, as it instead happens in the case of the k_t algorithm ($k = 1$); the Cambridge/Aachen (C/A) algorithm ($k = 0$) has energy-independent clustering. The anti- k_t algorithm is an infrared and collinear safe algorithm (IRC) for its distance definition. IRC safety indicates that the set of hard jets found remains unchanged even in case of a collinear splitting or the addition of a soft emitted gluon.

The reconstructed jets are calibrated through the ATLAS EM+JES scheme that applies corrections as a function of the jet energy and pseudo-rapidity to the jets reconstructed at the electromagnetic scale.

Due to multiple pp collisions within the same bunch crossing, a variety of particles not belonging to the primary interaction vertex are produced (*pile-up*). The *pile-up* products can interfere or by overlapping with the objects of physics interest, requiring an additional calibration correction, or generating new fake jets. The *pile-up* contribution to jet calibration is accomplished by subtracting the average additional energy due to *pile-up* interactions from the energy measured by the calorimeters. The

correction constants used for that purpose are obtained by *in situ* measurements coming from minimum bias data and are depends on the number of reconstructed primary vertices (NPV), the jet pseudo-rapidity (η) and the bunch spacing. Fake jets originated from *pile-up* fluctuations are rejected using the Jet Vertex Fraction (JVF) algorithm; JVF is defined as the fraction between the number of tracks p_T -matched to be originating from the primary interaction vertex and all the tracks from the hard scattering. The effect of the JVF application on 2011 data can be seen in Figure 4.1 where the number of jets becomes independent from the number of primary vertices (red distribution). *Pile-up* jets have very small JVF values as most of their tracks originate from additional *pile-up* vertices.

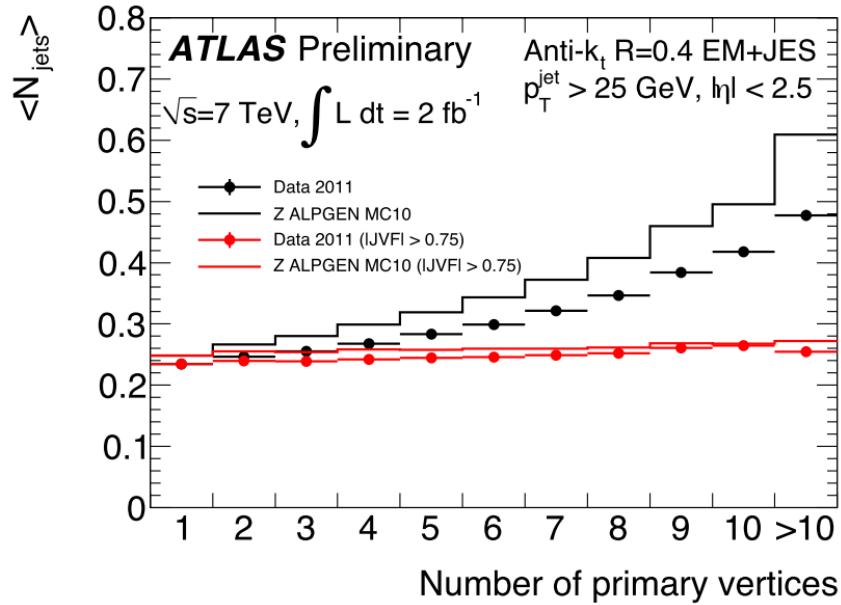


Figure 4.1 – Average number of jets reconstructed for increasing pile-up conditions, without (black) or with (red) Jet Vertex Fraction (JVF) cut applied [55][56]

The energy resolution for jets (JER) has been measured for data acquired in 2011 with both the di-jet balance [48] and the bisector [49] techniques, finding a good agreement with the MC resolution. For 2012 data, only the bisector method has been used and the results generally agree with the resolution obtained from MC, but small differences have been seen in some p_T and η ranges. The fractional jet p_T resolution with 2011 data for anti- k_t jets $R = 0.6$ is shown in Figure 4.2, comparing the two methods mentioned above.

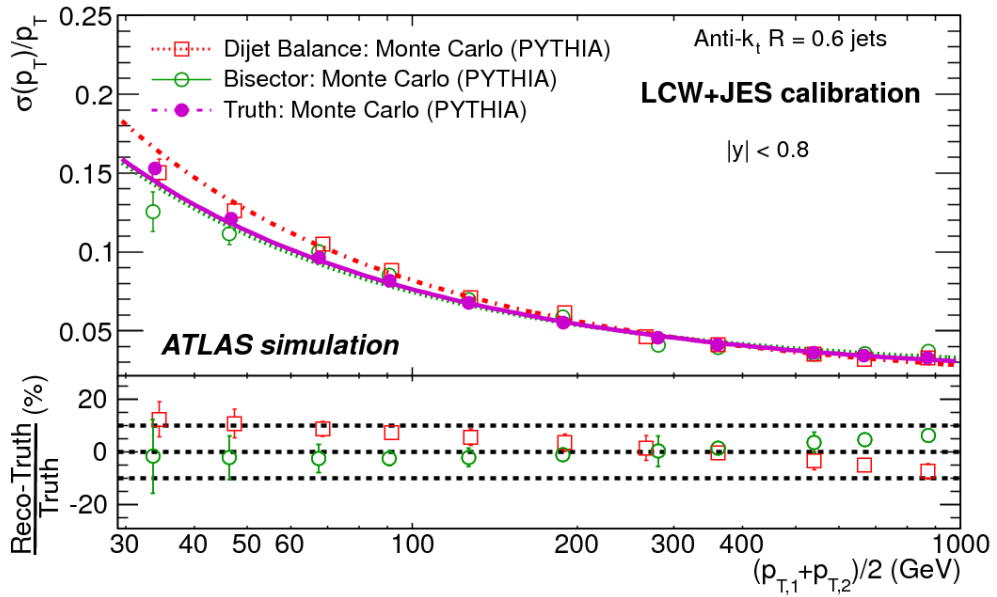


Figure 4.2 – Fractional jet p_T resolutions for anti- k_t jets with $R=0.6$ for the Local Cluster Weighting (LCW+JES), Global Cell Weighting Comparison between the Monte Carlo simulation truth jet p_T resolution and the final results obtained from the bisector and dijet balance in situ methods (applied to Monte Carlo simulation). The lower panels show the relative differences, obtained from the fits, between the in situ methods and Monte Carlo truth results. The dotted lines indicate relative differences of $\pm 10\%$. The errors shown are only statistical. (GCW+JES) and Global Sequential (GS) calibrations[75].

For the sake of these analyses it is very important to identify if a jet is originated by a b quark or a lighter one. This is done by specific *b-tagging* algorithms, which discriminates using some specific jet properties, in particular the wider impact parameter and the relatively long life-time of the b quarks, that is reflected in a measurable (few millimeters) flight length from the primary interaction vertex. In these analyses the MV1 *b-tagging* algorithm is used, based on a neural network strategy at a working point of 70% efficiency.

4.2 Electrons

The electron reconstruction performed in ATLAS is based on the matching between Inner Detector tracks and EM calorimeter clusters. The information from the EM calorimeter defines the energy of the electron while the ID tracks give the angular direction.

The ATLAS reconstruction algorithm is based on information coming from the different layers of the electromagnetic calorimeter, energy leakage in the hadronic one, track quality criteria from ID objects and cluster-track matching. The ATLAS recipe ensures a good discrimination from background objects by mainly requiring electron isolation (see further for isolation details). Three levels of electron identification called loose, medium and tight are defined, each with progressively more stringent requirements. The tight electrons are the ones used in the analyses and have to pass the following requirements:

- ❖ No error occurred in the Liquid Argon electromagnetic calorimeter during data taking;
- ❖ The track should be identified as “*tight++*” electron by a specific algorithm, based on the shape of the shower deposit in the calorimeter, the matching of the cluster to the associated track and the number of hits in the inner part of the tracker. The algorithm working point efficiencies is about 78% for electron in Z boson events;
- ❖ A small distance is required between the track impact parameter and the primary vertex projection on the z-axis, $|z_0^{PV}| < 2mm$;
- ❖ The transverse energy $E_T = \frac{E_{cluster}}{\cosh \eta_{track}}$ should be above a fixed threshold $E_T > 25GeV$;
- ❖ Compatibility with the calorimeter acceptance requirements $0 < |\eta_{cluster}| < 1.37$ and $1.52 < |\eta_{cluster}| < 2.47$;
- ❖ Isolation cuts on “*Etcone20*” and “*Ptcone30*” variables. The “*Etcone20*” is the total E_T deposited in the calorimeter towers in a cone of radius $\Delta R = 0.2$ around the electron position. “*Ptcone30*” is an analogue variable built by summing the p_T of the tracks in the ID around a cone of $\Delta R = 0.3$. The working points for both variable cuts are collected in a $\eta - p_T$ matrix, characterized by an efficiency of 90%.

The algorithm performance is robust with respect to increasing *pile-up*, as shown in Figure 4.3. An *in situ* calibration is used in ATLAS in order to fine tune the electromagnetic energy scale, providing a mass resolution of 1.6 GeV in $Z \rightarrow ee$ events, as shown in Figure 4.4. The well-known mass of the Z boson and its decay in

e^+e^- pairs are used to improve the knowledge of the electron energy scale and the linearity of the electromagnetic calorimeter.

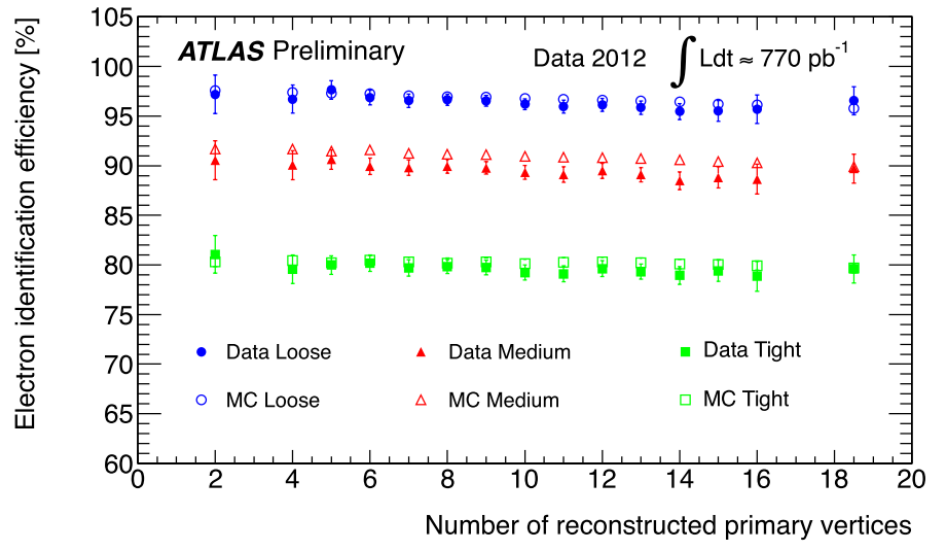


Figure 4.3 – Stability of electron identification efficiency with increasing pile-up, for different efficiency values [54]

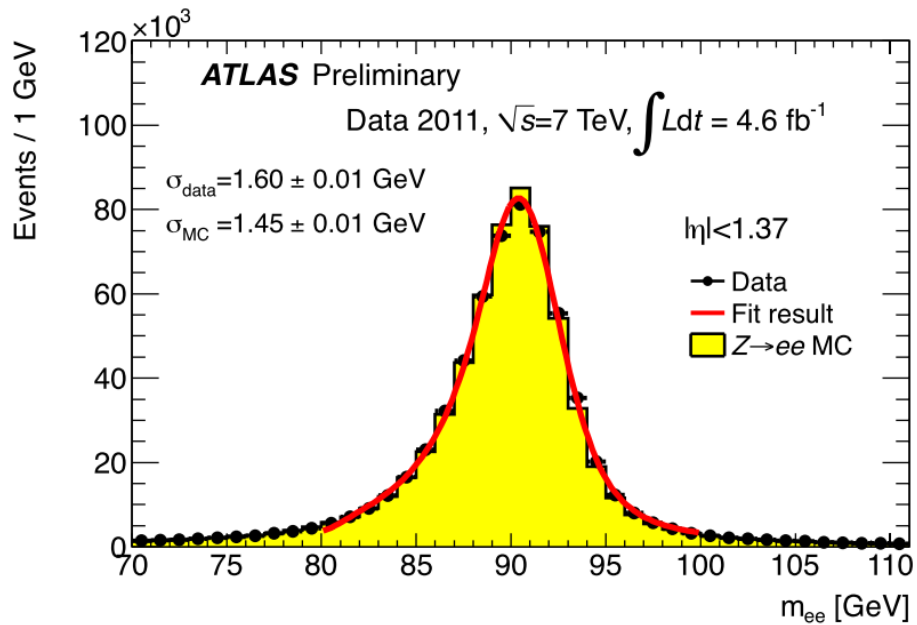


Figure 4.4 – Mass resolution in data and simulation for $Z \rightarrow ee$ events [54]

4.3 Muons

The muon reconstruction system uses information from ID tracks, calorimeter cells and Muon Spectrometer accomplishing with identification efficiency greater than 95% and a relative momentum resolution running from 3% to 10% depending on the muon p_T . The main reconstruction hit information are extracted from the MS system: three layers of precision drift tubes (MDT) chambers in $|\eta| < 2$ and two layers of the MDT chambers in combination with one layer of cathode strip chambers (CSC) at the beginning of the MS for $2 < |\eta| < 2.7$.

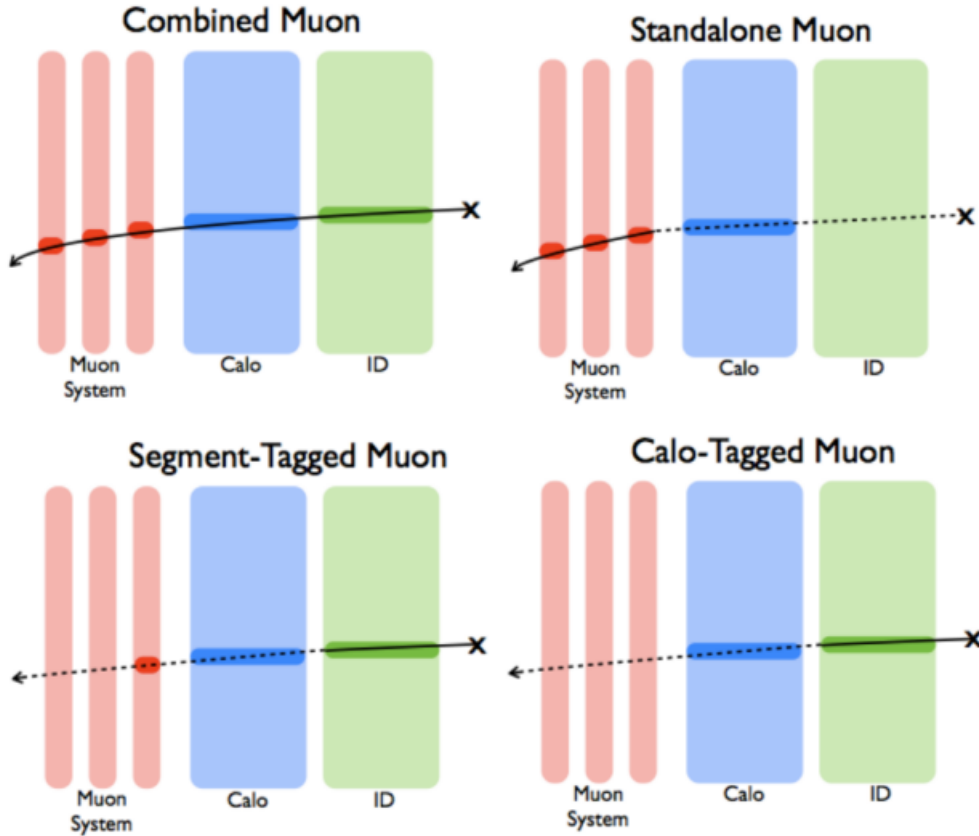


Figure 4.5 – The four kinds of reconstructed muon candidates in ATLAS: Combined muon require information from all the three detectors, Standalone muons need MS identification, Segment-tagged looks a muon.

In ATLAS, four kinds of muon candidates (*combined*, *standalone*, *segment-tagged* and *calo-tagged*) are distinguished depending on detector information used for their identification as shown in Figure 4.5. Two reconstruction algorithms are used in ATLAS, *STACO* [50] and *MuId* [51]; they differ only in the combination scheme of ID and MS tracks that is respectively done as a statistical combination of all track vectors or as a refitting procedure starting from the ID tracks to the MS ones. In the specific analyses presented in this thesis the *MuID* muons have been used, satisfying the following requirements:

- ❖ To be identified as a “*tight*” muon, that imply to be a *combined muon* or a *standalone muon* with at least three MDT+CSC hits;
- ❖ The projection on the beam direction of the impact parameter of the muon track with respect to the primary vertex should be small $|z_0^{PV}| < 2mm$;
- ❖ Isolation requirement: $Etcone20 < 4GeV$ and $Ptcone30 < 2.5 GeV$

The muon reconstruction efficiency as a function of pile-up is shown in Figure 4.6 for combined candidates with the additional isolation requirement. Simulations well reproduce the behavior observed in data showing a high robustness against the effects of pile-up (Figure 4.7).

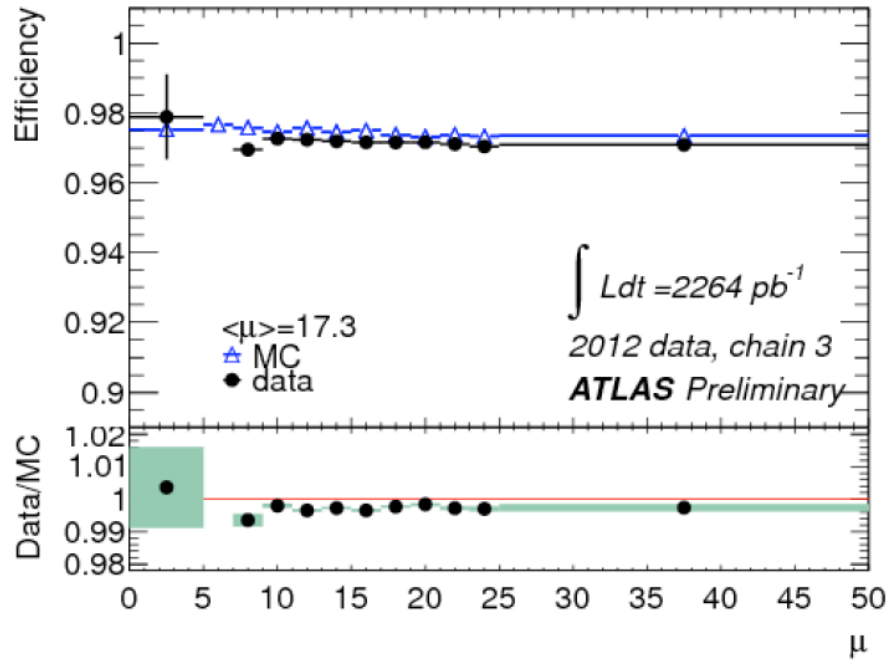


Figure 4.6 – Stability of muon identification efficiency [53]

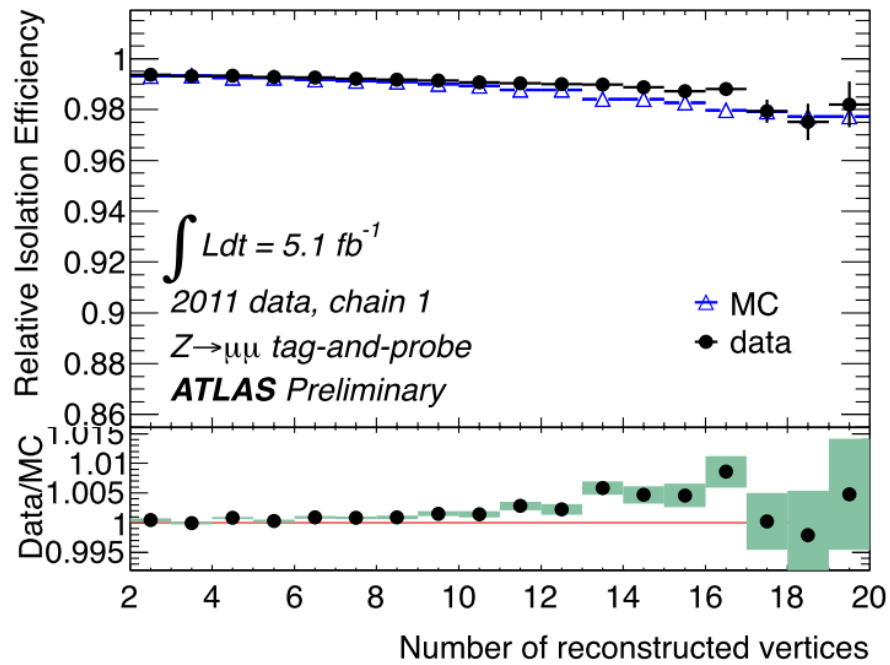


Figure 4.7 – Stability of muon isolation efficiency with increasing pile-up, for combined muons reconstructed in both the inner detector and muon spectrometer [53]

4.4 Missing Transverse Energy

In hadron colliders, the initial momentum of the colliding partons along the beam axis is not known a priori, so the amount of total missing energy cannot be determined. Anyway, the sum of the interacting partons transverse momentum is in good approximation zero with respect to the beam axis, allowing the measurement of the event missing transverse energy (E_T^{miss}) defined as

$$E_T^{miss} = \sqrt{(E_x^{miss})^2 + (E_y^{miss})^2} \quad 4.3$$

The ATLAS algorithm for the E_T^{miss} evaluation includes contributions from topo-clusters transverse energy, corrections for energy losses in the cryostat system and reconstructed muons[52]

$$E_{x,tot}^{miss} = E_{x,calo}^{miss} + E_{x,cryo}^{miss} + E_{x,\mu}^{miss} \quad 4.4$$

where the calorimeter term $E_{x,calo}^{miss}$ is evaluate only from cells belonging to topological clusters and included in the pseudo-rapidity range $|\eta| < 4.9$. Each calorimetric cell is calibrated in a different way depending on the reconstructed objet it belongs to. Indeed the $E_{x,calo}^{miss}$ is the sum of different components evaluated as the negative sum of the energy deposit in the calorimetric cells associated to the correspondent object typology

$$E_{x,calo}^{miss} = E_{x,e}^{miss} + E_{x,\gamma}^{miss} + E_{x,\tau}^{miss} + E_{x,jet}^{miss} + E_{x,\mu(calor)}^{miss} + E_{x,CellOut}^{miss} \quad 4.5$$

where $E_{x,CellOut}^{miss}$ is evaluated from the topo-cluster cells not included in none of the reconstructed objects.

The cryostat term $E_{x,cryo}^{miss}$ get rid of the non-negligible loss of energy in hadronic showers due to the cryostat system installed between the LAr electromagnetic calorimeter and the Tile hadronic calorimeter. It is evaluated via the energy correlation between the last LAr layer and the first Tile one. The muon term $E_{x,\mu}^{miss}$ is evaluated from the ID and MS muon information. Analog considerations and definitions are for the component along the y axis.

The E_T^{miss} performance and systematic uncertainties are established from differences between data and simulations distribution in $Z \rightarrow \ell\ell$ and $W \rightarrow \ell\nu$ events, as shown in Figure 4.8. Figure 4.9 shows the resolution of the E_T^{miss} in $Z \rightarrow \mu\mu$ events as a function of the number of primary vertices before and after the pile-up suppression, in data and simulations.

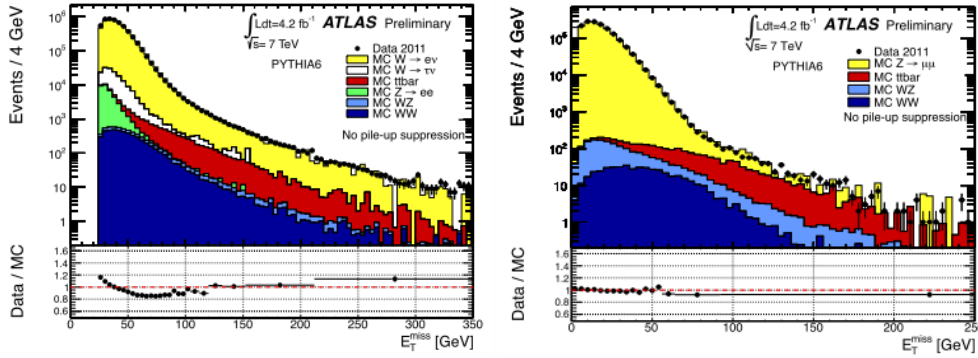


Figure 4.8 – MET in events without neutrinos ($Z \rightarrow \mu\mu$, left) and with neutrinos ($W \rightarrow \mu\nu$, right) [52]

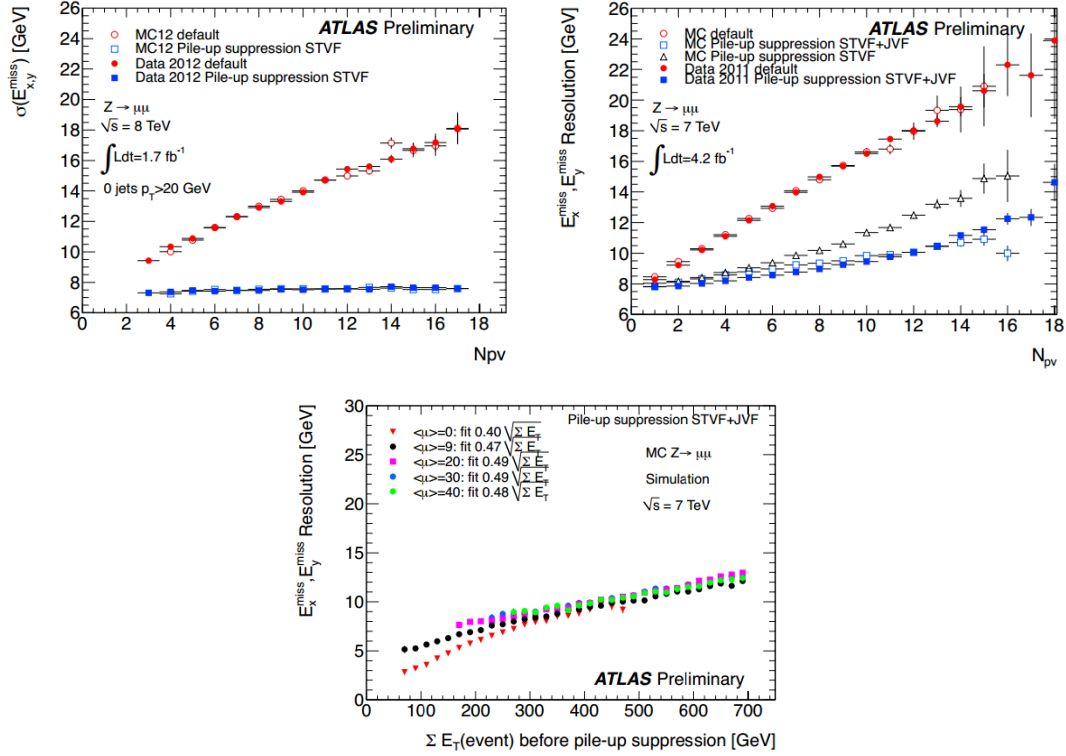


Figure 4.9 – Left: MET resolution, Middle: E_x^{miss}, E_y^{miss} resolution, vs number of primary vertices in the event (N_{pv}). Right: pile-up corrected E_x^{miss}, E_y^{miss} resolution vs Sum E_T , for different values of the average number of interactions per bunch crossing (μ) [52].

4.5 Boosted Objects

The LHC is exploring a completely new physics regime where the available center-of-mass energy far exceeds the masses of known standard model particles. At such energies, heavy particles such as W and Z bosons and top quarks are often produced with large transverse momentum (*boosted particles*) that implies large Lorentz boost for their decay products. As a consequence, final state products are close enough to

make traditional reconstruction algorithms significantly less efficient. In the 2012 data analysis presented in this thesis, the large integrated luminosity collected at $\sqrt{s} = 8 \text{ TeV}$ allows to explore the high- p_T region for $t\bar{t}$ events with unprecedented sensitivity. At this time, only the cross section dependence with respect to hadronic top has being studied.

The decay products of hadronically decaying top quarks are largely overlapping and can be reconstructed as a single, energetic and large radius jet (*fat jet* or *large- R jet*). In Figure 4.10 it is possible to see a graphical representation of the jet configuration coming from a low- p_T top quark decay (on the left) with respect to one coming from a high- p_T top quark (in the center); on the right, the last configuration is shown how it appears in a single, *large- R jet* reconstruction. In Figure 4.11 the average distance ΔR between the quarks produced in a top decay process is plotted as a function of the top momentum; from this distribution it is possible to appreciate the jet approaching dependence at the growing of the top p_T . Already from a $p_T^t \approx 500 \text{ GeV}$ overlapping problems start, in fact the average jet distance ΔR is smaller than the double of the anti- k_t cone radius parameter usually fixed at 0.4 in standard jet reconstruction.

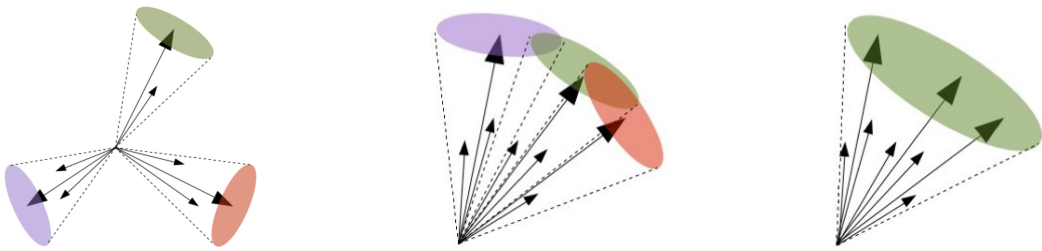


Figure 4.10 – Graphical representation of the jet produced in top quark event decays, in case of low (left) and high (center) values of top p_T . The picture on the right show the same high top p_T configuration as before using a large- R jet reconstruction.

A single fat jet, that contains all the decay products of a massive particle, will have significantly different properties than a jet of the same p_T originated from a light-quark or gluon. The characteristic heavy particle decay signature results in a hard substructure that is absent in the light-quark and gluon jets. In the latest few years new techniques, called *tagging algorithms*, have been proposed in order to recognize *fat jets* originated by massive particle decays with the aim to increase efficiency and purity in high energy analysis. Such techniques involve the study of the substructure of fat jets both via direct comparison of the decay signature and by using a selection based on substructure variables as jet mass, splitting scale and N-subjettiness [62].

Tagging algorithms can be more effective, especially in substructure variables selection, by removing soft radiation coming from fat jets. The selective removal of soft radiation is performed in many ways and it is generally referred to jet “*grooming*” methods.

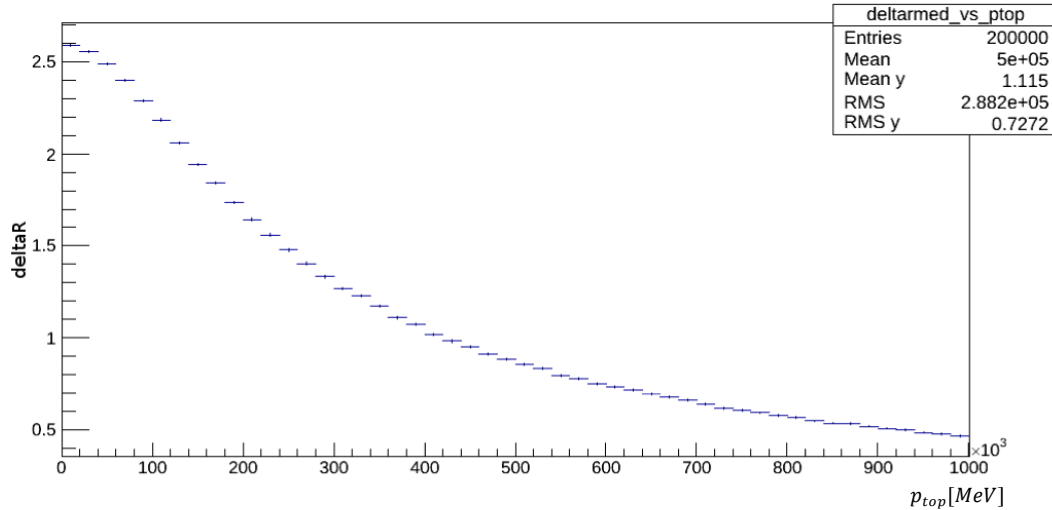


Figure 4.11 – Distribution of the average distance between the produced by a top quark decay as a function of the top p_T itself. Events from the Powheg+Pythia generator used to simulate $t\bar{t}$ signal events in the 2012 data analysis are used.

In the following paragraphs an overview of the most useful quantities for *top tagging* and the principals grooming techniques used in ATLAS are presented. In Paragraph 4.8 HepTopTagger algorithm is presented, while in Paragraph 4.9 the Top Template Tagger method is described.

4.6 Jets substructure observables

The principal jet substructure variables used by ATLAS tagging algorithms are jet mass, splitting scale and N-subjettiness, described in the following.

JET MASS m_{jet} is calculated from the energies and momenta of the constituents of the jet, as given in the equation

$$m_{jet} = \sqrt{\left(\sum_i E_i\right)^2 - \left(\sum_i p_i\right)^2} \quad 4.6$$

where E_i and p_i are respectively the energy and momentum's absolute value of the i^{th} jet obtained summing up its constituents, namely the *topo-clusters* and the tracks. The jet mass is a very important quantity searching for boosted, high-mass particles; it can be used as a powerful discriminant between signal and background.

SPLITTING SCALES variables are evaluated during k_t -algorithms reconstruction steps as the k_t -distance of the two jets. The splitting scale variable is defined as

$$\sqrt{d_{ij}} = \min(p_{T,i}, p_{T,j}) \cdot \Delta R_{ij} \quad 4.7$$

where ΔR_{ij} is the geometric distance between the two sub-jets. The most useful spitting scales for boosted tagging are the one measured at the last reclustering step, namely the $\sqrt{d_{12}}$ observable, and the $\sqrt{d_{23}}$ variable defined at the second to the last step of the reclustering.

N-SUBJETTINESS variables τ_N are observables related to the sub-jet multiplicity. The τ_N variable is calculated by clustering the constituents of the jet requiring exactly N sub-jets to be found; in order to evaluate this quantities, only the k_t algorithm is used. The k_t algorithm, in this case, interrupts the clusterization process when there are exactly N *pseudo-jets* remaining. The variables τ_N are then defined as the sum over all k constituents of the jet:

$$\tau_N = \frac{1}{d_0} \sum_k p_{T_k} \times \min(\delta R_{1k}, \delta R_{2k}, \dots, \delta R_{Nk}) \quad 4.8$$

$$d_0 \equiv \sum_k p_{T_k} \times R$$

where R is the jet radius parameter, p_{T_k} is the p_T of the k^{th} constituent and δR_{ik} is the distance between the i^{th} sub-jet to k^{th} constituent. From this definition, τ_N indicates how well the fat jet can be described as containing N or fewer subjets, discriminating by how constituents are localized close to the sub-jet axes. In order to discriminate a fat jet derived from a boosted top quark with respect to one originated by the parton shower of a light quarks or a gluons, the ratio τ_3/τ_2 (τ_{32}) can be used. [63,64]

4.7 Jet grooming algorithms

In this section three of the most diffused grooming algorithms used in ATLAS are presented.

MASS-DROP FILTERING procedure identifies relatively symmetric sub jets, each with a mass significantly smaller than the one of the whole fat jet; only the three most energetic sub-jets are conserved. This technique was developed to be used with C/A jets reconstruction and is optimized for the Higgs boson searches in the two b quarks decaying channel $H \rightarrow b\bar{b}$ [65]. For that reason, it is not so effective on top tagging on its own but remains an important starting point for other techniques.

TRIMMING algorithm [66] removes contamination from pile-up, multiple parton interactions (MPI), and initial-state radiation (ISR) that are often much softer than hard-scattering partons products. The selection criteria used is based on the p_T ratio of the jet constituents. The trimming procedure uses k_t algorithm to reconstruct smaller sub-jets from the *fat jet* constituents, putting on them the p_T constraint $p_{T,i}/p_{T,fatjet} < f_{cut}$, where $p_{T,i}$ is the transverse momentum of the i^{th} sub jet, and f_{cut} is a parameter of the method; typical values of f_{cut} are around a few percent. The surviving constituents form the *trimmed jet*. This procedure is illustrated in Figure 4.12. Low-mass jets from a light quark or gluon usually lose 30–50% of their mass, while jets containing the decay products of a boosted object lose only a few percent of mass, mainly removing pileup contribution; this is due to the fat jet internal structure that is more uniform in the case of light quarks and gluons production.

PRUNING algorithm [67, 68] is similar to trimming because it removes soft constituents from the fat jet, but it adds a wide-angle radiation veto. The fat jet constituents are used to reconstruct again the jet, using either a C/A or k_t algorithm; at each pseudo jet recombination step the following pruning cuts are placed

$$\frac{p_T^{j_2}}{p_T^{j_1+j_2}} > z_{cut}$$

$$\Delta R_{j_1,j_2} < R_{cut} \cdot \frac{2m^{jet}}{p_T^{jet}} \quad 4.9$$

where j_1, j_2 are the pseudo jet considered in the current step ordered $p_T^{j_1} > p_T^{j_2}$, R_{cut} and z_{cut} are parameters of the tagger. It is important to remark that these requirements are not directly related to the original fat jet but to the proto-jets formed in the new reconstruction process.

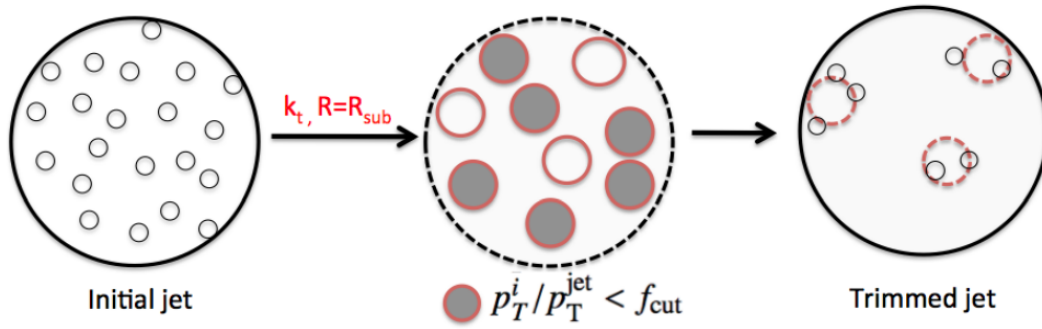


Figure 4.12 – Pictorial explanation of the jet trimming procedure

In Figures from 4.13 to 4.16 the effect of the trimming algorithm on distributions of mass, splitting scales and N-subjettiness is shown, referred to the leading- p_T jet in the range $600 \leq p_T^{fatjet} < 800$ GeV. For these studies a $Z' \rightarrow t\bar{t}$ Monte Carlo sample ($m_{Z'} = 1.6$ TeV) has been considered for signal-like events (red lines), compared with a MC multijets background (black lines). The effect of grooming increases the separation between signal and Monte Carlo distributions for all the substructure variables considered, helping the discrimination based on these quantities. The application of trimming and splitting scale variable $\sqrt{d_{12}}$ selection in a boosted based analysis will be treated deeper in Chapter 5.

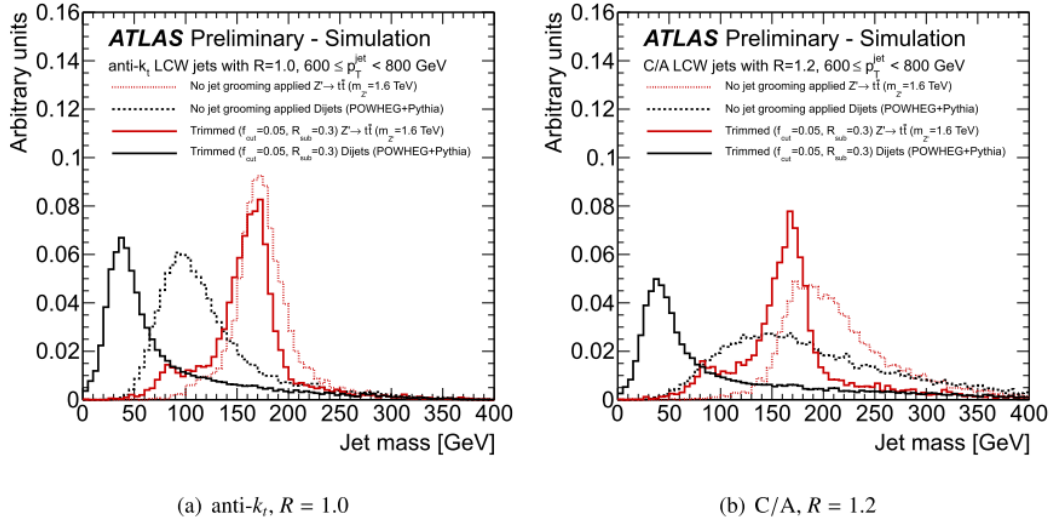


Figure 4.13 – Leading- p_T jet mass comparing $Z' \rightarrow t\bar{t}$ ($m_{Z'} = 1.6$ TeV) signal to POWHEG multi-jet background for jets in the range $600 \leq p_T^{jet} < 800$ GeV. The dotted lines show the ungroomed leading- p_T jet distribution, while the solid lines show the corresponding trimmed ($f_{cut} = 0.05, R_{sub} = 0.3$) jets. The groomed distributions are normalized with respect to the ungroomed distributions, which are themselves normalized to unity [96].

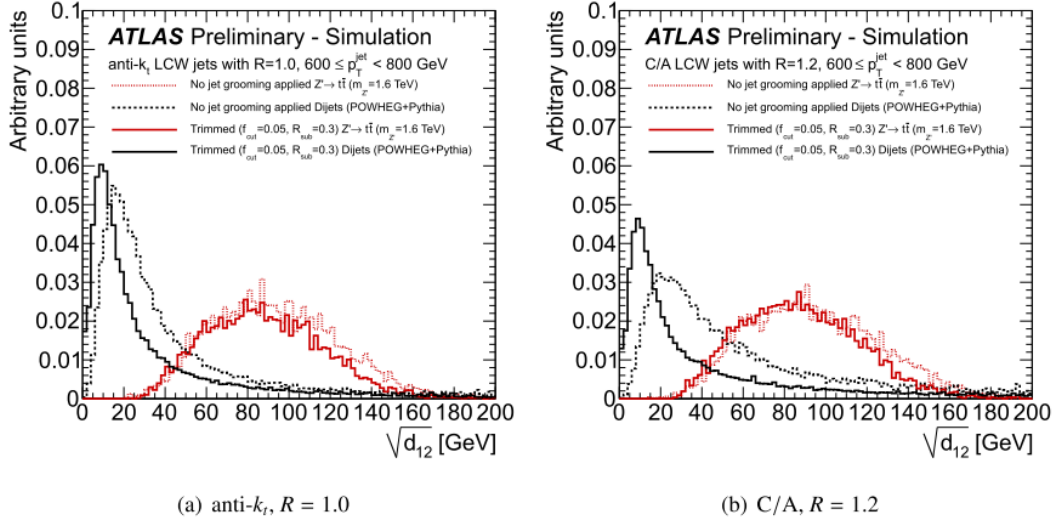


Figure 4.14 – Leading- p_T jet splitting scale $\sqrt{d_{12}}$ comparing $Z' \rightarrow t\bar{t}$ ($m_{Z'} = 1.6$ TeV) signal to POWHEG multi-jet background for jets in the range $600 \leq p_T^{\text{jet}} < 800$ GeV. The dotted lines show the ungroomed leading- p_T jet distribution, while the solid lines show the corresponding trimmed ($f_{\text{cut}} = 0.05, R_{\text{sub}} = 0.3$) jets. The groomed distributions are normalized with respect to the ungroomed distributions, which are themselves normalized to unity [96].

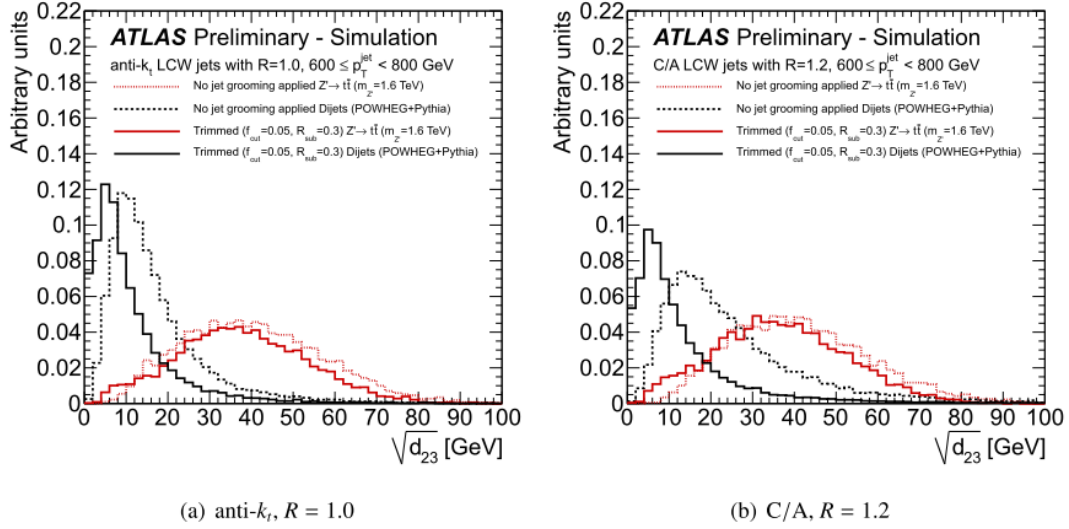


Figure 4.15 – Leading- p_T jet splitting scale $\sqrt{d_{23}}$ comparing $Z' \rightarrow t\bar{t}$ ($m_{Z'} = 1.6$ TeV) signal to POWHEG multi-jet background for jets in the range $600 \leq p_T^{\text{jet}} < 800$ GeV. The dotted lines show the ungroomed leading- p_T jet distribution, while the solid lines show the corresponding trimmed ($f_{\text{cut}} = 0.05, R_{\text{sub}} = 0.3$) jets. The groomed distributions are normalized with respect to the ungroomed distributions, which are themselves normalized to unity [96].

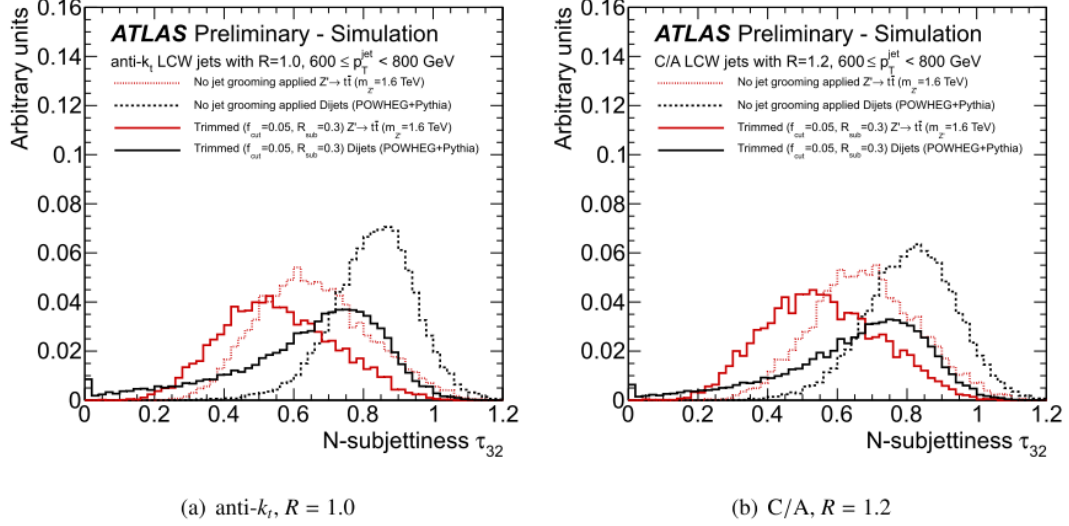


Figure 4.16 – Leading- p_T jet N-subjettiness comparing $Z' \rightarrow t\bar{t}$ ($m_{Z'} = 1.6$ TeV) signal to POWHEG multi-jet background for jets in the range $600 \leq p_T^{\text{jet}} < 800$ GeV. The dotted lines show the ungroomed leading- p_T jet distribution, while the solid lines show the corresponding trimmed ($f_{\text{cut}} = 0.05, R_{\text{sub}} = 0.3$) jets. The groomed distributions are normalized with respect to the ungroomed distributions, which are themselves normalized to unity [96].

4.8 HEPTopTagger

HEPTopTagger is an example of how the jet grooming techniques may be used in order to optimize the selection of the hadronically decaying boosted top quarks [69]. The method uses a variant of the mass-drop filtering technique on C/A jet algorithm. The HEPTopTagger chain is described in the following and graphically exemplified in Figure 4.17.

- The two sub-jets j_i in the last clustering stage of the C/A algorithm have to satisfy the criteria

$$\frac{m^{j_i}}{m^{jet}} < \mu_f \quad 4.10$$

where μ_f is a parameter of the tagger. This procedure is back-iterative applied to all sub-jets passing the mass cut, until m^{j_i} is less than a fixed mass parameter m_{cut} . At least three subjects must survive or the fatjet will be rejected;

- The combinations of three substructure objects are filtered at a time, reconstructing their constituents using the C/A algorithm with the radius parameter $R = \min \left[0.3, \frac{\Delta R_{j_1, j_2}}{2} \right]$, where $\Delta R_{j_1, j_2}$ is the minimum separation between all possible pairs in the current triplet.
- The sum of the resulting n protojets should be near to the mass of the top quark ($140 \leq m^{jet} < 200 \text{ GeV}$);
- Only a number $N_s > 3$ of these n protojets are then again reconstructed to form a new jet triplet that will be properly calibrated [24];
- The triple is accepted as a top candidate if one of the following criteria on the invariant mass of two (m_{ij}) and three (m_{123}) subjects combinations is satisfied.

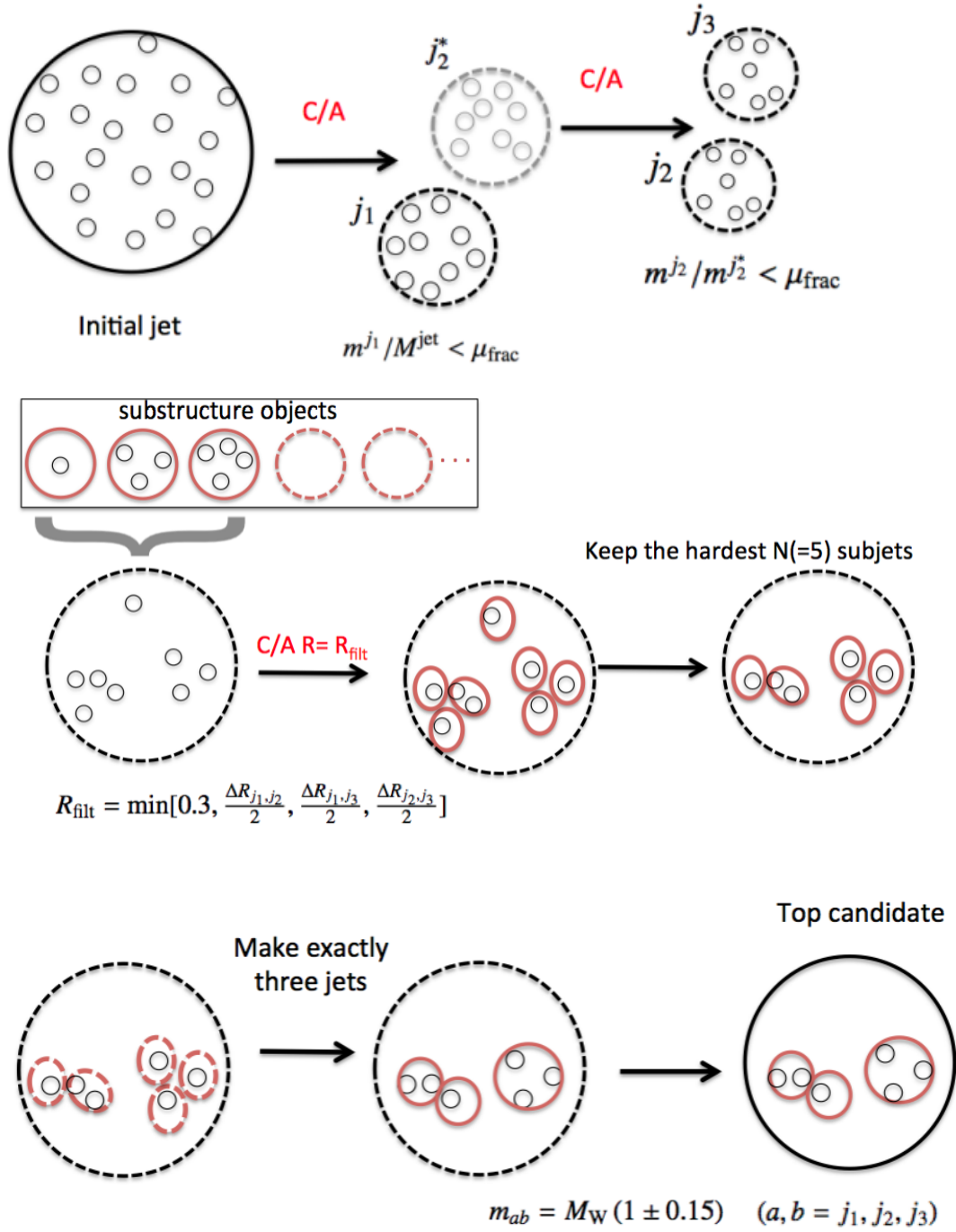


Figure 4.17 – Pictorial explanation of the different steps in the HEPTopTagger algorithm.

$$\begin{aligned}
0.2 < \arctan \frac{m_{13}}{m_{12}} < 1.3 \quad \text{and} \quad R_{min} < \arctan \frac{m_{23}}{m_{123}} < R_{max} \\
R_{min}^2 \left(1 + \left(\frac{m_{13}}{m_{12}} \right)^2 \right) < 1 - \left(\frac{m_{23}}{m_{123}} \right)^2 \\
&< R_{max}^2 \left(1 + \left(\frac{m_{13}}{m_{12}} \right)^2 \right) \quad \text{and} \quad \frac{m_{23}}{m_{123}} > 0.35 \\
R_{min}^2 \left(1 + \left(\frac{m_{12}}{m_{13}} \right)^2 \right) < 1 - \left(\frac{m_{23}}{m_{123}} \right)^2 \\
&< R_{max}^2 \left(1 + \left(\frac{m_{12}}{m_{13}} \right)^2 \right) \quad \text{and} \quad \frac{m_{23}}{m_{123}} > 0.35
\end{aligned} \tag{4.11}$$

and R_{min} and R_{max} are method parameters.

4.9 Top Template Tagger

The Top Template Tagger method [70, 71] discriminates fat jets containing top quark products with respect to various background sources by quantifying, through an infra-red safe estimator, how well the fat jet overlaps with a simulated set of top decay hypothesis (the *template*). The overlap estimator, ranging from 0 to 1, quantifies the agreement in *energy flow* between simulated top quark decays and observed jets. The *energy flow* is to be intended as the energy distribution of all the jet clusters on a $\eta - \varphi$ surface [72].

A huge amount of templates simulating many top quark decay configurations needs to be generated for different values of top transverse momentum in order to cover with a sufficient granularity the full accessible phase space. A single template event is

formed by three four-vectors representing the three partons produced in a top quark decay: one prompt b quark and two originated from the W boson decay. The top quark is simulated only along one direction and the template triplet will then be rotated event by event to match the fat jet axis.

The overlap estimator Ov_3 is defined as a function of the difference between the energy of each simulated parton E_i and the sum of the energy of the single *fat jet topo-clusters* E_{cl} that are close enough to the parton axis. It is defined as

$$Ov_3 = \max_{\{\tau_n\}} \exp \left[- \sum_{i=1}^3 \frac{1}{2\sigma_i^2} \left(E_i - \sum_{\Delta R < \Delta R_{limit}} E_{cl} \right)^2 \right] \quad 4.12$$

where $\{\tau_n\}$ is the set of templates and ΔR is the $\eta - \varphi$ distance if the i^{th} parton and the topo-cluster. ΔR_{limit} can be fixed to a certain radius (usually 0.2) or variable with respect to the top p_T as follow

$$\Delta R_{limit}^{variable} = - \frac{A}{p_{T,fatjet}^2} + \frac{B}{p_{T,fatjet}} + C \quad 4.13$$

where A, B and C parameters are calculated from Monte Carlo training. The weighting variable has been evaluated $\sigma_i = E_i/3$ from performance studies, judged looking at tagging efficiency and background rejection.

The Template Overlap rejection power has been tested comparing simulation of $t\bar{t}$ signal and multijets background events, generated with different MC generators. The resulting Ov_3 overlap variable distributions, represented in Figure 4.18, show a

completely different behavior between top, in blue, and QCD multi-jets, in pink, events; the superposition of the two distribution is represented in violet. The different behavior of the two distributions allows a quite good discrimination between top quark signal and QCD multi-jets background.

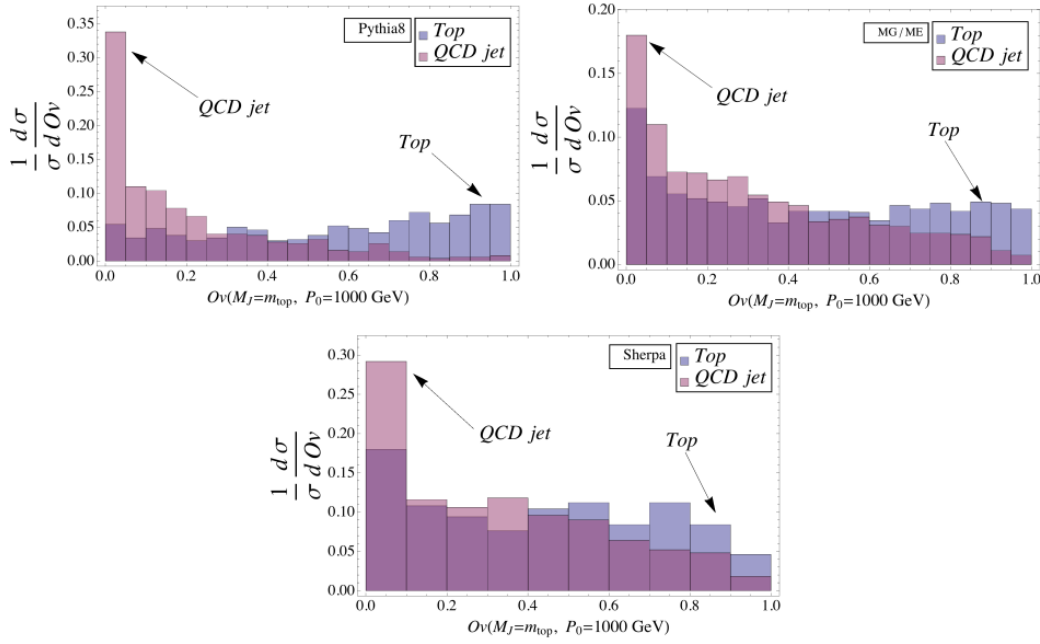


Figure 4.18 – Comparison of histograms of template overlap discrimination variable Ov_3 for top quark and QCD jets from different MCs [upper left (right) Pythia (MG/ME) and Sherpa on the bottom].

4.10 TOM efficiency and rejection power

The application of a sophisticated *top-tagger* as the TOM algorithm instead of the simple cut-based selection on jet substructure variables now used, is the natural follow up for the 2012 analysis described in this thesis. Some preliminary efficiencies and

rejection studies will be presented in this paragraph in order to test the algorithm response when applied to simulated events. Different physical processes have taken into account: SM $t\bar{t}$ production, Z' boson decaying to high momentum $t\bar{t}$ pairs, W +jets and QCD multi-jets backgrounds.

The events used for these studies have passed the selection described in Paragraph 5.1.2 but for the requirements of at least one b-tagged jet and at least a *large- R jet* fulfilling the kinematic cuts. The TOM algorithm is then applied to the leading p_T *non-trimmed large- R jet* in each event. The efficiency has been evaluated choosing a selection cut at $Ob_3 = 0.7$

$$\varepsilon = \frac{n_{events}(Ob_3 > 0.7)}{n_{events}(tot)} \quad 4.14$$

while the background rejection power has been defined as

$$r = 1 - \varepsilon \quad 4.15$$

Z' BOSON SAMPLE

Monte Carlo events of very massive ($m_{Z'} = 2TeV$) Z' boson production decaying to $t\bar{t}$ pair are particularly convenient in order to study the Ob_3 distribution of high p_T top quarks. Thanks to the high Z' mass, top quarks are produced with a much higher momentum than the ones from SM direct production. In that way the abundance of high p_T top is larger and the elaboration time consuming in the generation phase is shorter. For the Z' generation the *ALPGEN* simulator is used. In Figure 4.19 the distribution of the

Ov_3 overlap estimator is shown; the event distribution clearly peaks close to $Ov_3 \sim 1$, implying a good matching. Dividing the jet p_T spectrum in 6 subsamples (*bins*) and producing the Ov_3 distributions for each one of them, it is clear from Figure 4.20 that the jets with a small Ov_3 value are concentrated in the low p_T bins. The same and more quantitative conclusion can be seen calculating the tagging efficiency for each p_T bin in case $Ov_3 > 0.7$, as shown in Figure 4.21. From these results it is possible to assert that the TOM algorithm tagging efficiency grows with the growing of the jet p_T .

DIRECT $t\bar{t}$ PRODUCTION FROM SM PROCESSES

The same study has been repeated using a SM $t\bar{t}$ MC sample generated with *Powheg+Pythia*, as the one used in the 2012 data analysis, in order to simulate the signal events. The Ov_3 discriminant distribution shown in Figure 4.22 is peaked at a value close to zero, on the contrary to what happens in the Z' case; however, looking at the jet p_T -bin distributions in Figure 4.23, it is easy to recognize in high p_T regions a similar behavior as in the Z' case, with the distributions peaked to 1 only for high jet p_T . The opposite result in Figure 4.24 is due to the great number of events coming from the relatively low p_T bins, where the TOM is less performing and the Ov_3 distributions peak to 0. The efficiency as a function of the jet p_T is shown in Figure 4.24, reaching a maximum value of ~ 0.64 .

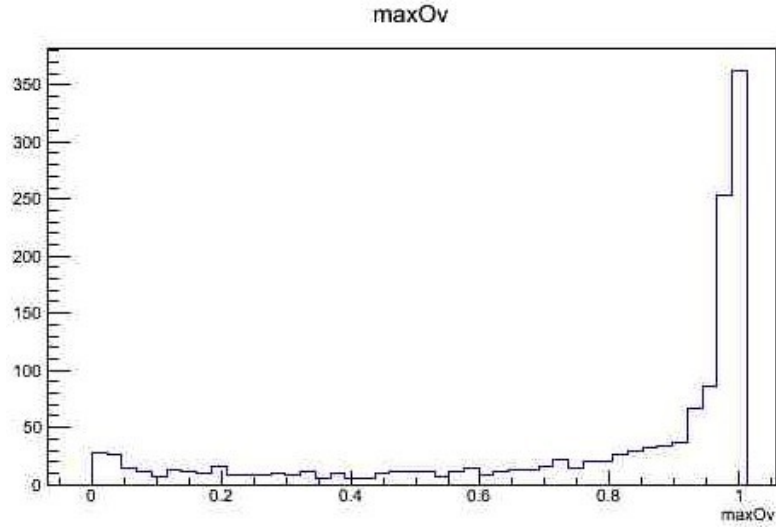


Figure 4.19 – TOM overlap estimator Ov_3 distribution of $Z' \rightarrow t\bar{t}$ events. A Ov_3 value close to 1 means a good event matching, while the matching is worst getting close to 0.

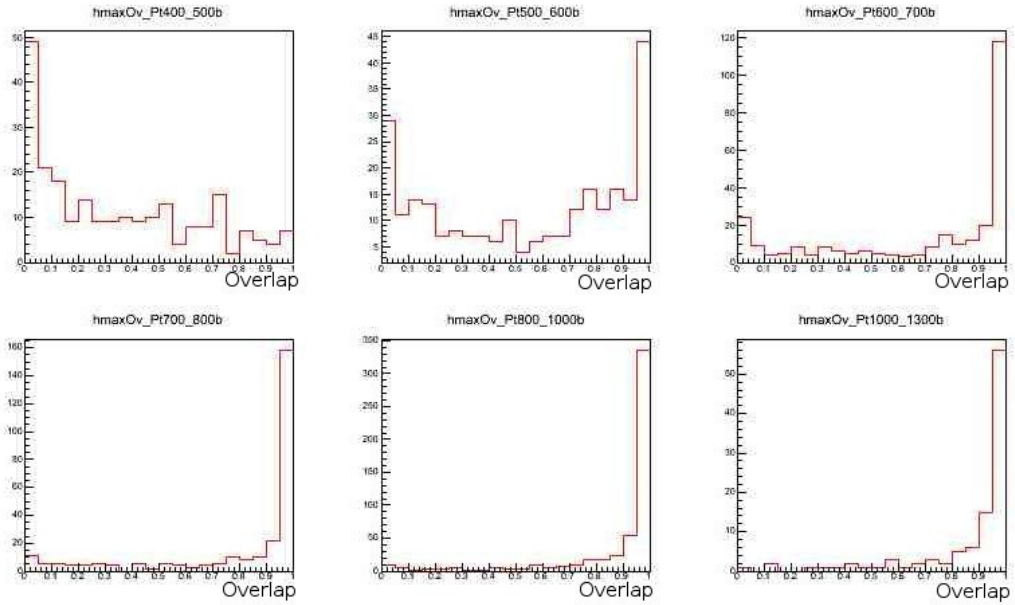


Figure 4.20 – Overlap estimator Ov_3 distributions of $Z' \rightarrow t\bar{t}$ events divided in hadronic top p_T subsample. It is possible to see how the Top Template tagger performances becomes better at increasing top quark p_T values.

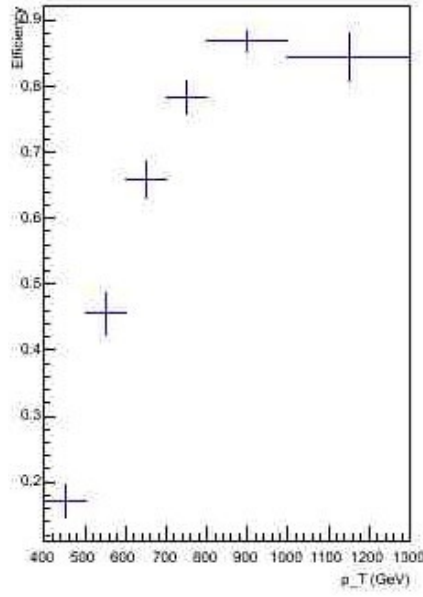


Figure 4.21 – Distribution of the TOM tagger efficiency for each hadronic top p_T bin in which the $Z' \rightarrow t\bar{t}$ sample has been divided; tagging efficiency strongly grows with p_T^t .

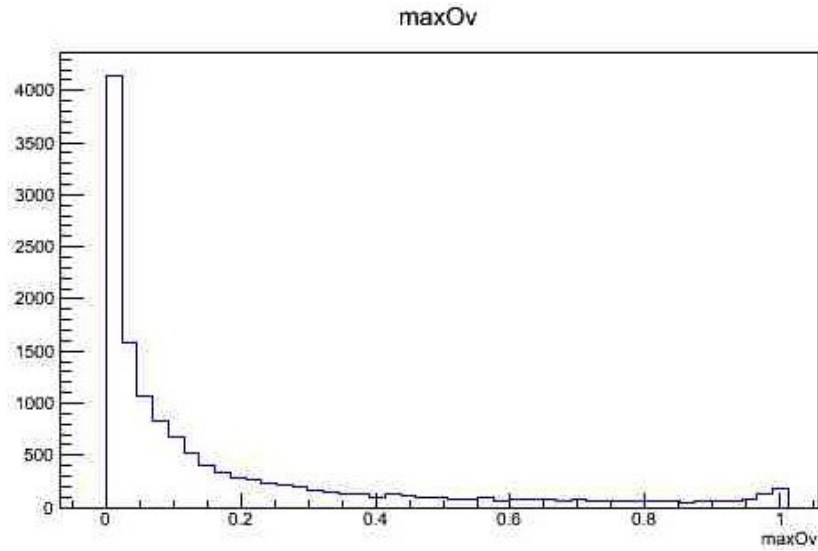


Figure 4.22 – TOM overlap estimator Ov_3 distribution of SM production $t\bar{t}$ events. A Ov_3 value close to 1 means a good event matching, while the matching is worst getting close to 0.

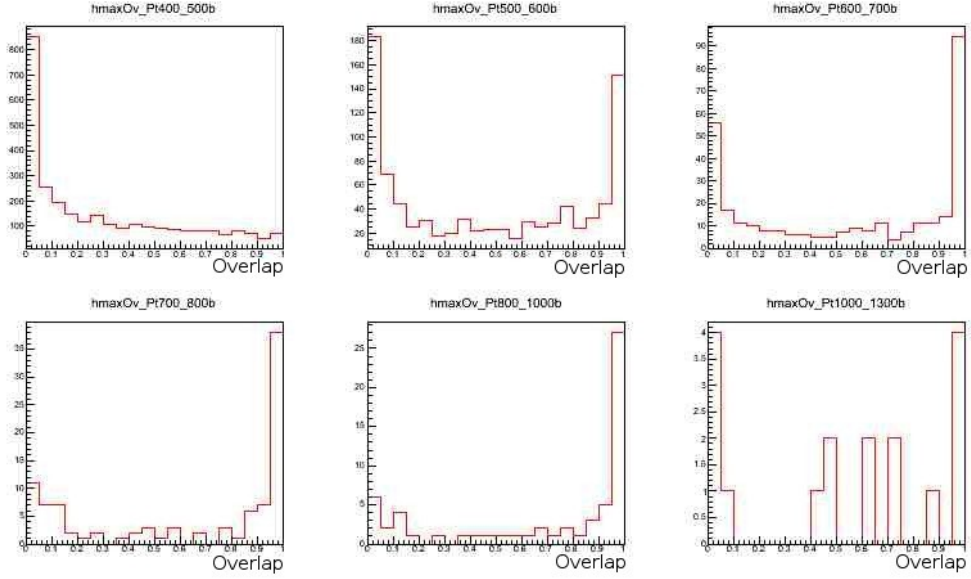


Figure 4.23 – Overlap estimator Ov_3 distributions of SM production $t\bar{t}$ events divided in hadronic top p_T subsample. It is possible to see how the Top Template tagger performances becomes better at increasing top quark p_T values. The last bin has little significance due to the lack of statistic in the MC at such high p_T^t .

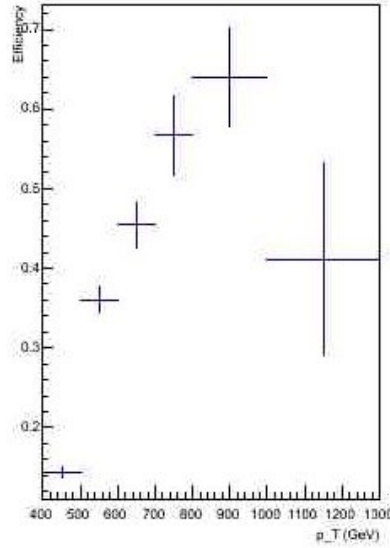


Figure 4.24 – Distribution of the TOM tagger efficiency for each hadronic top p_T bin in which the $t\bar{t}$ sample has been divided; tagging efficiency strongly grows with p_T^t .

***W* + *JETS* BACKGROUND**

Applying the TOM on the $W + jets$ background used in the 2012 data analysis, the resulting Ov_3 distributions are peaked at zero in every jet p_T bin, as shown in Figures 4.25 and 4.26. This confirms the expectation of low matching in case of non- $t\bar{t}$ events. The efficiencies also are lower with respect to signal cases, never exceeding the value $\varepsilon = 0.25$, corresponding to a rejection power always greater than $r = 0.75$.

QCD MULTI-JETS BACKGROUND

Similar considerations as for the $W + jets$ channel can be done in the case of the QCD multi-jet MC background sample simulated with ALPGEN. The corresponding plots are shown in Figures 4.27 and 4.28; results are slightly worst than in the $W+jets$ case, leading to a maximum efficiency of $\varepsilon \sim 0.35$, that means a minimum rejection power $r \sim 0.65$.

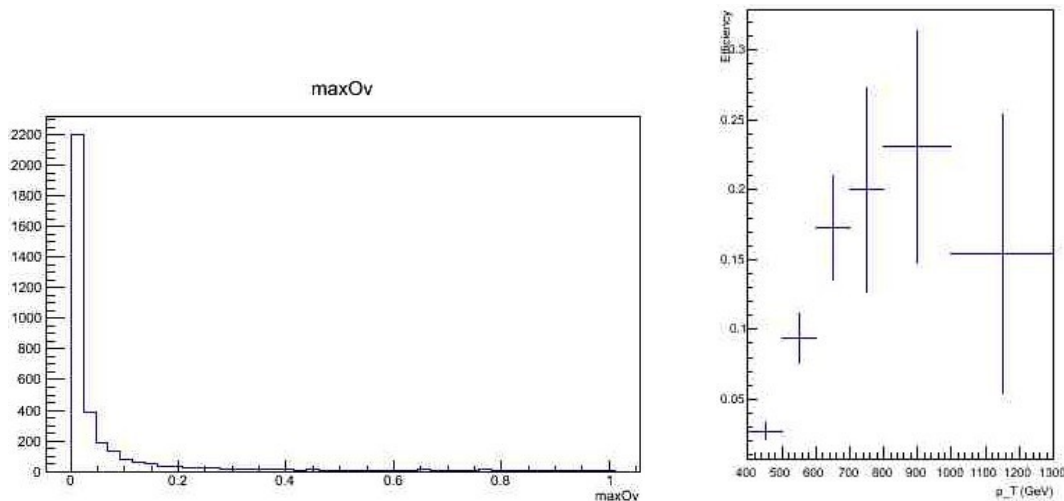


Figure 4.25 – TOM overlap estimator Ov_3 distribution of $W+jets$ events (left) and distribution of tagging efficiency for each hadronic top p_T bin in which the sample has been divided (right).

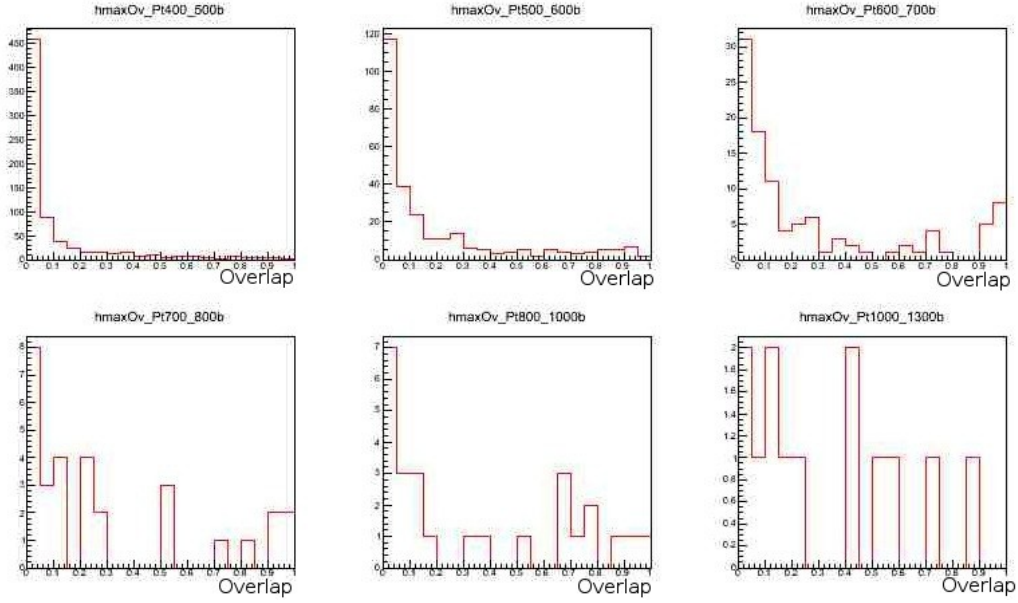


Figure 4.26 – Overlap estimator Ov_3 distributions of W +jets events divided in hadronic top p_T subsamples. The Ov_3 value remains at low values in each plot as expected for background events.

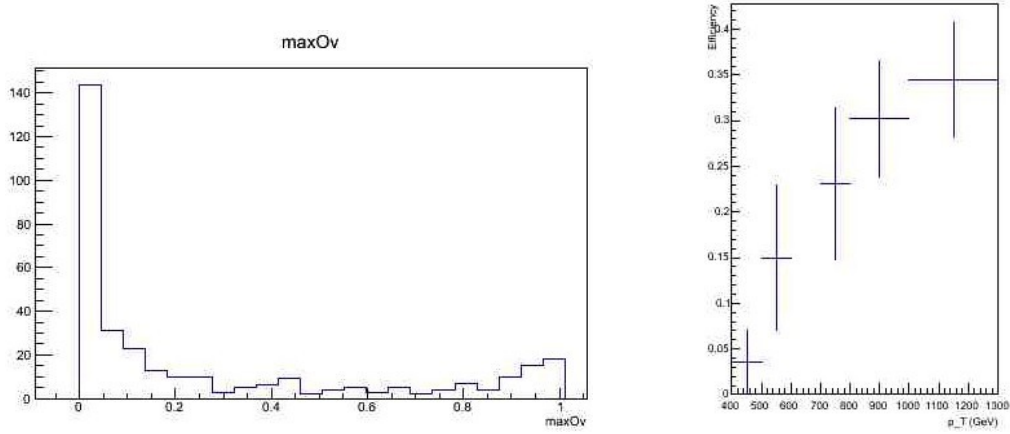


Figure 4.27 – TOM overlap estimator Ov_3 distribution of QCD multi-jets events (left) and distribution of tagging efficiency for each hadronic top p_T bin in which the sample has been divided (right).

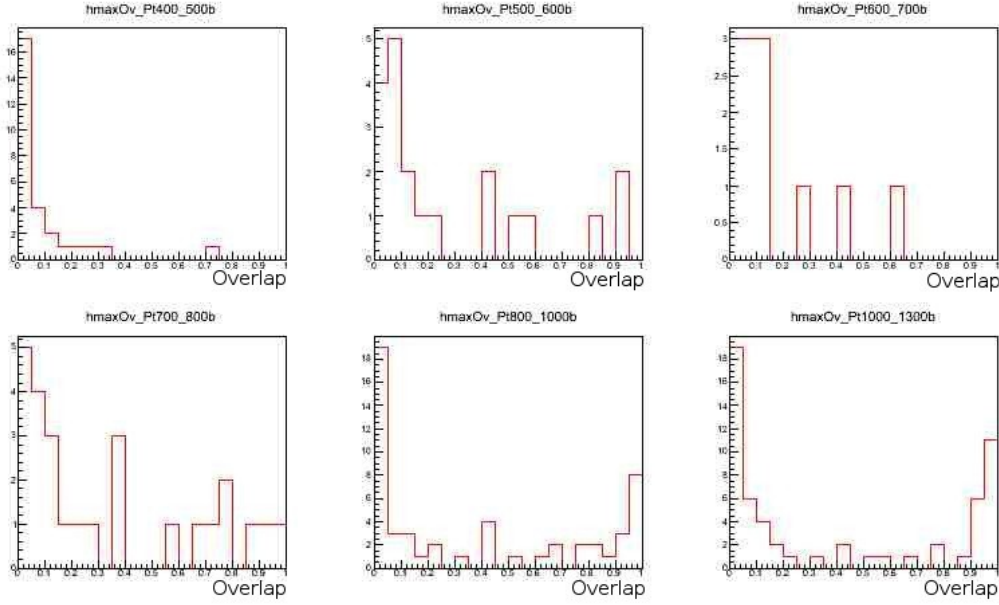


Figure 4.28 – Overlap estimator Ov_3 distributions of QCD multi-jets events divided in hadronic top p_T subsamples. The Ov_3 value remains at low values in each plot as expected for background events.

In Table 4.1 the number of events passing the TOM selection $Ov_3 > 0.7$ is reported for each MC sample considered and compared to the number of events before the selection and to the ones passing the full analysis selection described in Paragraph 5.1.2. While the results for the Z' sample are comparable in the two cases, for both $t\bar{t}$ signal and backgrounds, the cut-based selection gives better results. This result can be explained by a not optimized tuning of the algorithm parameters, such as a fixed ΔR_{limit} value instead of a more performing p_T -variable one; the choice has been done in order to be more conservative and comparable with previous obtained results. Moreover the TOM has been designed for very high p_T phase space regions, as the ones that could be

reached during the next LHC data taking period at $\sqrt{s} = 13 \text{ TeV}$ and simulated with enough statistic only by Z' MC samples. In this case the algorithm is more efficient as it can be seen in the Z' case. Much work is still ongoing to improve the TOM tuning, both for the present analyses and with a view on the future data taking.

Sample	events before tagging	Standard Analysis	TOM [$ Ov_3$]
Z	1366	1088	1099
$t\bar{t}$	13515	16687	12955
W+jets	3509	57	79
QCD	333	24	60

Table 4.1 – Summary of the number of events for each sample used in the Top Template tagging studies.

Chapter 5

Analysis Method and Event Selection

Two different analyses are considered in this thesis; for both of them the event selection used will be described in Paragraph 5.1, the description of the physical objects reconstruction will be done in Paragraph 5.2 and the comparison between the measured and the simulated events will be shown in Paragraph 5.3. In Paragraph 5.4 the unfolding procedure used will be discussed in detail while the description of the structure of the analysis code developed and used in the 2012 data analysis is accomplished in Paragraph 5.5.

The first analysis exploits the full 2011 data set collected at a center of mass energy of $\sqrt{s} = 7 \text{ TeV}$, with an integrated luminosity of $\int \mathcal{L} dt = 4.66 \text{ fb}^{-1}$, in order to

measure the differential production cross section of $t\bar{t}$ pairs in the $\ell + jets$ channel with respect to different kinematic variables of the top quark and the $t\bar{t}$ system: transverse momentum ($p_T^t, p_T^{t\bar{t}}$), mass ($m^{t\bar{t}}$) and rapidity ($y^{t\bar{t}}$). In the SM, $t\bar{t}$ events are produced predominantly at rest, i.e. the $p_T^{t\bar{t}}$ distribution is strongly peaked at zero, and the produced top quarks are limited in a p_T^t range of a few hundreds GeV . The event selection is therefore tuned for relatively low p_T top quarks, whose decay products are spatially well separated, so this kinematic phase space is called *resolved regime* [76].

The second analysis aims to study the high p_T^t region with higher efficiency, selecting events where the top quarks decay products are subject to strong Lorentz boost effects (*boosted regime*), and then strongly collimated, leading to event signatures that cannot be efficiently identified with the standard algorithms used in *resolved* analyses. These events are characterized by a large overlap among jets coming from the hadronically decaying top quark that can be contained in a single jet with a larger radius parameter (usually denoted as *large- R jet* or *fat jet*). The purpose of this analysis is therefore the measurements of the differential production cross section of $t\bar{t}$ pairs in the $\ell + jets$ channel with respect to the hadronic top quark p_T in the high transverse momentum region; the full 2012 data set collected by the ATLAS detector ($\int \mathcal{L} dt = 20 fb^{-1}$) at a center of mass energy of $\sqrt{s} = 8 TeV$ has been used [77].

In both the analyses $t\bar{t}$ events are discriminated by a proper selection scheme as described in Paragraph 5.1 while the appropriate top quark and $t\bar{t}$ kinematic variables are evaluated as respectively described in Paragraph 5.2. In order to compare the measured events with the theoretical predictions, an unfolding procedure is performed to take into account distortions introduced by the limited resolution and acceptance of the detector. In the following the event distributions after the unfolding application will

be referred to as “*unfolded*”. In the boosted analysis case, the cross section result is compared with theoretical predictions only in a fiducial region of the phase space, in order to reduce the errors due to the data extrapolation to the fully acceptance phase space. Unfolding strategy and fiducial level definition are described further in this chapter.

5.1 Event Selection

After the initial trigger selection described in Chapter 3, additional requirements are applied to the event sample in order to enhance the purity of the selected $t\bar{t}$ candidates. The event selection cuts in $e + jets$ and $\mu + jets$ channels are similar each other, leading to similar yields and distributions, therefore allowing an easier combination of the electron and muon channels.

5.1.1 Resolved Analysis

The event selection cuts applied are the following:

- 1) Events must belong to the so called *good run list* of events acquired when all detectors work properly;
- 2) A reconstructed primary vertex with five or more associated tracks and no electromagnetic or hadronic calorimeter corrupted data;
- 3) Exactly one good reconstructed electron/muon with $p_T > 25\text{GeV}$ matching with the triggered one;

- 4) No good reconstructed leptons of different flavour with $p_T > 15 \text{ GeV}$;
- 5) At least four good reconstructed jets and none bad reconstructed jets in the events;
- 6) $E_T^{miss} > 30 \text{ GeV}$;
- 7) $m_T^W > 35 \text{ GeV}$. It is computed from the reconstructed lepton and the E_T^{miss} , defined as $m_T^W = \sqrt{2p_T^\ell p_T^\nu - \cos(\varphi^\ell - \varphi^\nu)}$;
- 8) At least one *b-tagged* jet.

The choice of requiring just one *b-tagged* jet instead of two is due to the decision to maintain a larger efficiency and to avoid a further systematic uncertainties increasing, that would be caused by the poor knowledge of the b-tag correction scale factor.

Jets are tagged as “*good jets*” if they pass some general quality criteria (dealing with hardware problems, cosmic rays, beam-gas interactions, and so on) and a *jet vertex fraction* request of $|JVF| < 0.75$, that enhance the fraction of jets coming from the hard scattering. The additional kinematic cuts of $p_T > 25 \text{ GeV}$ and $|\eta| < 2.5$ are also applied.

5.1.2 Boosted Analysis

In the boosted top analysis two types of jets have been considered: *small-R jets*, reconstructed by fixing the R parameter in the anti- k_T algorithm equal to 0.4, are candidates for the b quark coming from the leptonic quark decay; *large-R jets*, with the

anti-kt algorithm parameter $R = 1$, includes the whole hadronic top quark decay products. The $t\bar{t}$ event candidates must succeed all the following requirements:

- 1) Events must belong to the so called *good run list* of events acquired when all detectors work properly;
- 2) A reconstructed primary vertex with five or more associated tracks and no LAr or Tile calorimeter corrupted data;
- 3) Exactly one good reconstructed lepton candidate per event matching with the triggered one;
- 4) No good reconstructed leptons of different flavour;
- 5) E_T^{miss} greater than 30GeV (electron channel) or 20 GeV (muon channel);
- 6) $m_T^W > 30\text{GeV}$ in case of electrons. In the muon channel a so called “*triangular cut*” is required on E_T^{miss} and the transverse mass of the W boson: $E_T^{miss} + m_T^W > 60\text{ GeV}$;
- 7) At least one *small-R* jet candidate close to the lepton, $\Delta R(jet, lepton) < 1.5$;
- 8) At least one *large-R* jet with $p_T > 300\text{ GeV}$, $m > 100\text{ GeV}$ and the splitting scale $\sqrt{d_{12}} > 40\text{ GeV}$, after the trimming algorithm applied;
- 9) At least one *b-tagged small-R jet*.

5.2 Physical objects definition

After the $t\bar{t}$ candidates' selection, it is necessary to define the physical objects and their kinematical quantities; in the following paragraphs the different identification methods are presented.

5.2.1 Resolved Analysis

In order to evaluate the kinematic variables of the top quarks and of the $t\bar{t}$ pairs, the full $t\bar{t}$ system is reconstructed using *KLFitter* [57], a kinematic fit based on a likelihood approach which assesses the compatibility of the observed events with the expected $t\bar{t}$ decays.

The input to the likelihood fit are: the measured energy, the pseudo-rapidity and the azimuthal angle of four jets ($\hat{E}_{jet,i}, \hat{\eta}_{jet,i}, \hat{\phi}_{jet,i}$), the measured lepton energy (\hat{E}_ℓ), and the missing transverse momentum assumed to be due to a single neutrino ($\hat{p}_{T,\nu}$). The likelihood is maximised with respect to the partons' energy ($\tilde{E}_{part,i}$), the lepton transverse momentum ($\tilde{p}_{T,\ell}$) and the three momentum components of the neutrino (\tilde{p}_X). The $t\bar{t}$ likelihood for $\ell + jets$ events is expressed as the product three components:

- A product of Breit-Wigner \mathfrak{B} distributions dealing with the distribution probability of W bosons and top quarks production. These Breit-Wigner functions depend on the partons and leptons theoretical predictions and need as input the W boson and top quark pole masses and decay widths, respectively fixed to the measured values: $m_W = 80.4 \text{ GeV}$, $m_t = 172.5 \text{ GeV}$, $\Gamma_W = 2.1 \text{ GeV}$ and $\Gamma_t = 1.5 \text{ GeV}$;

- A product of Transfer Functions (\mathbb{T}) allowing to match the reconstructed quantities (\hat{X}) to quarks and leptons produced in the hard scattering (\tilde{X}). These transfer functions are derived from MC@NLO simulation, with all corrections applied. A truth mapping was used to separate *b-jets* from light jets; a specific set of transfer functions parameters exists for each type of jet;
- The *b*-tagging probability P evaluated from Monte Carlo simulation depending on the flavour of the parton.

The resulting likelihood is defined as follow:

$$\begin{aligned}
 \mathcal{L} = & \mathfrak{B}(\tilde{E}_{part,1}, \tilde{E}_{part,2} | m_W, \Gamma_W) \cdot \mathfrak{B}(\tilde{E}_\ell, \tilde{E}_\nu | m_W, \Gamma_W) \cdot \\
 & \cdot \mathfrak{B}(\tilde{E}_{part,1}, \tilde{E}_{part,2}, \tilde{E}_{part,3} | m_t, \Gamma_t) \cdot \\
 & \cdot \mathfrak{B}(\tilde{E}_\ell, \tilde{E}_\nu, \tilde{E}_{part,4} | m_t, \Gamma_t) \cdot \mathbb{T}(\hat{E}_x^{miss} | \tilde{p}_{x,\nu}) \cdot \\
 & \cdot \mathbb{T}(\hat{E}_y^{miss} | \tilde{p}_{y,\nu}) \cdot \mathbb{T}(\hat{E}_\ell | \tilde{E}_\ell) \cdot \prod_{i=1}^4 \mathbb{T}(\hat{E}_{jet,i} | \tilde{E}_{part,i}) \cdot \\
 & \cdot P(btag|quark)
 \end{aligned} \tag{5.1}$$

The likelihood maximization assigns four measured jets to the decay partons of the $t\bar{t}$ pair, taking into account all possible permutations of the five leading jets in the event. In order to further enhance the fraction of properly reconstructed $t\bar{t}$ pairs, an additional selection cut on the best permutation likelihood value \mathcal{L} is applied, requiring $\log \mathcal{L} > -50$. If no jet combination satisfies this request, the event is rejected. In Figure 5.1 the $\log \mathcal{L}$ distributions for $e + jets$ and $\mu + jets$ channels are reported.

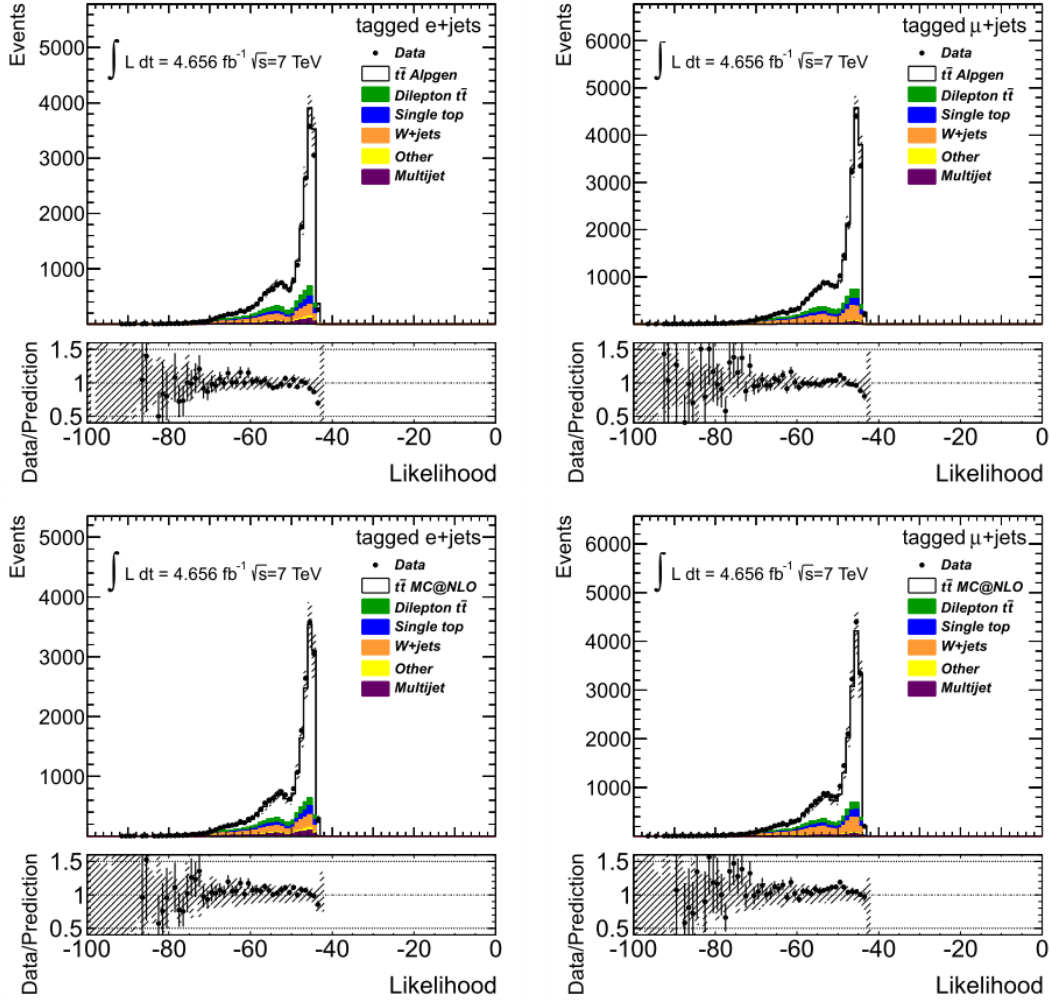


Figure 5.1 – The kinematic likelihood fit $\log \mathcal{L}$ distributions for the $e+jets$ channel (left) and $\mu+jets$ channel (right). All statistical and systematic uncertainties are taken into account in the error bands. Signal model: Alpgen (top) and MC@NLO (bottom).

5.2.2 Boosted Analysis

The decaying products from the boosted hadronic top quark are considered to be enough collimated to be contained in a single *large- R* jet that can be considered, in first

approximation, a good representation of the top quark itself. In this analysis *trimmed large- R* jets are considered, in order to enhance the top matching efficiency and to decrease *pile-up* contribution [66].

In this analysis, the jet *top-tagging* is performed exploiting the jet mass and k_t splitting scale $\sqrt{d_{12}}$ obtained in the last step of the jet clustering process performed by the anti- k_t algorithm. It is defined as

$$\sqrt{d_{12}} = \min(p_{T,1}, p_{T,2}) \cdot \Delta R_{12} \quad 5.2$$

where 1 and 2 are referred to the two pseudo-jets (or *sub-jets*) in the final step of the k_t reconstruction. This definition is equivalent to the square root of the k_t distance parameter in k_t algorithm defined in Equation 4.1 multiplied by the jet radius parameter R in order to remove the explicit dependence from the nominal jet radius. The splitting scale $\sqrt{d_{12}}$ is particularly effective in distinguishing heavy particle decays, which tend to be reasonably symmetric, from largely asymmetric splitting expected for light quark or gluon jets. The expected $\sqrt{d_{12}}$ value for a heavy particle case is approximately $\sqrt{d_{12}} \approx m_{jet}/2$ while in the light particle case its value steeply falls, as shown in Figure 5.2.

In the case of a large p_T , leptonically-decaying top, the reduction of the angles between the decay products due to the Lorentz boost can lead leptons to be produced close to the b -quark so that the efficiency for usual isolation criteria significantly decrease. In order to maintain high efficiency in the boosted region, the requirement on the lepton-jet distance has been changed from the generic $\Delta R_{(lepton,jet)} > 0.4$ to $\Delta R_{(electron,jet)} > 0.2$ and $\Delta R_{(muon,jet)} > 0.1$ respectively for electron and muon channels. The usual isolation criteria on calorimeter cells and tracks should be loosened

as well, maintaining a sufficient rejection rate from the non-prompt leptons coming from QCD jets. The adopted solution is the *mini isolation* discriminant [58], which exploits the fact that the angular separation between the decay products is proportional to the mass of the originating particle over its momentum m_x/p_x . *Mini-isolation* collects energy around the lepton with a variable cone size, which decreases as the momentum of the lepton increases. As the mass of top quark is much heavier than other quarks and gluons, the prompt lepton from boosted top quarks will generally be more separated from jets of similar momentum.

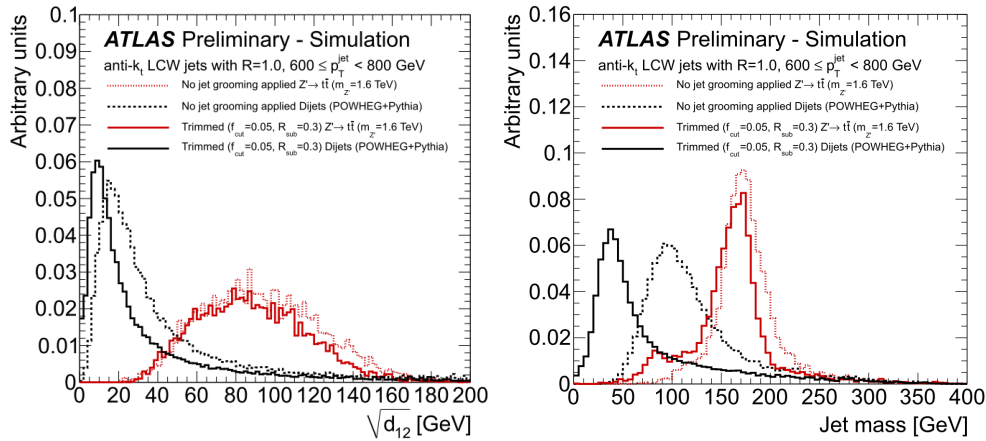


Figure 5.2 – Leading- p_T jet mass and splitting scale d_{12} comparing $Z \rightarrow q\bar{q}$ events to POWHEG multi-jet background for jets in the range $600 \leq p_T^{\text{jets}} < 800$ GeV. The dotted lines show the ungroomed jet distributions, while the solid lines show the trimmed. The trimmed parameters are $f_{\text{cut}} = 0.05$ and $R_{\text{sub}} = 0.3$. The groomed distributions are normalized with respect to the ungroomed distributions, which are themselves normalized to unity.[96]

5.3 Data/MC comparison plots

A manifold of plots of the principals quantities of physics interest have been produced to check that Monte Carlo simulations are in good agreement with the detector results in both electron and muon channels. In the following plots (from Figure 5.3 to Figure 5.7) real data are represented by black dots while MC contributions are differently colored depending on their $t\bar{t}$ signal or background origin, and stacked to be comparable with data. The binning has been decided in order to fill each bin with about the same number of events; this choice has been done to uniform statistical uncertainties among all the bins. The number of events plotted has been normalized to the corresponding bin width in order to avoid flat distributions. In plots referring to the 2011 analysis, the uncertainty band includes the systematic uncertainties as described in Chapter 6. In the plots from 2012 analysis the uncertainty bands represents only statistic uncertainty; the reason of this choice is explained in Chapter 6.

In the resolved analysis with 2011 data, the distributions are in general well described by the Monte Carlo simulation, as shown by the ratio present in the bottom part of each plot from Figure 5.3 to Figure 5.5.

In the boosted analysis a discrepancy on normalization of the order of 20–30% is observed between the data and MC in all the kinematic variable distributions considered. The origin of this normalization disagreement is still under investigation by comparing the results obtained by other analysis groups using similar requirements and by looking at the effects caused by variations in the event selection. Data/MC comparisons for some relevant kinematic variables are shown in Figures 5.6 and 5.7.

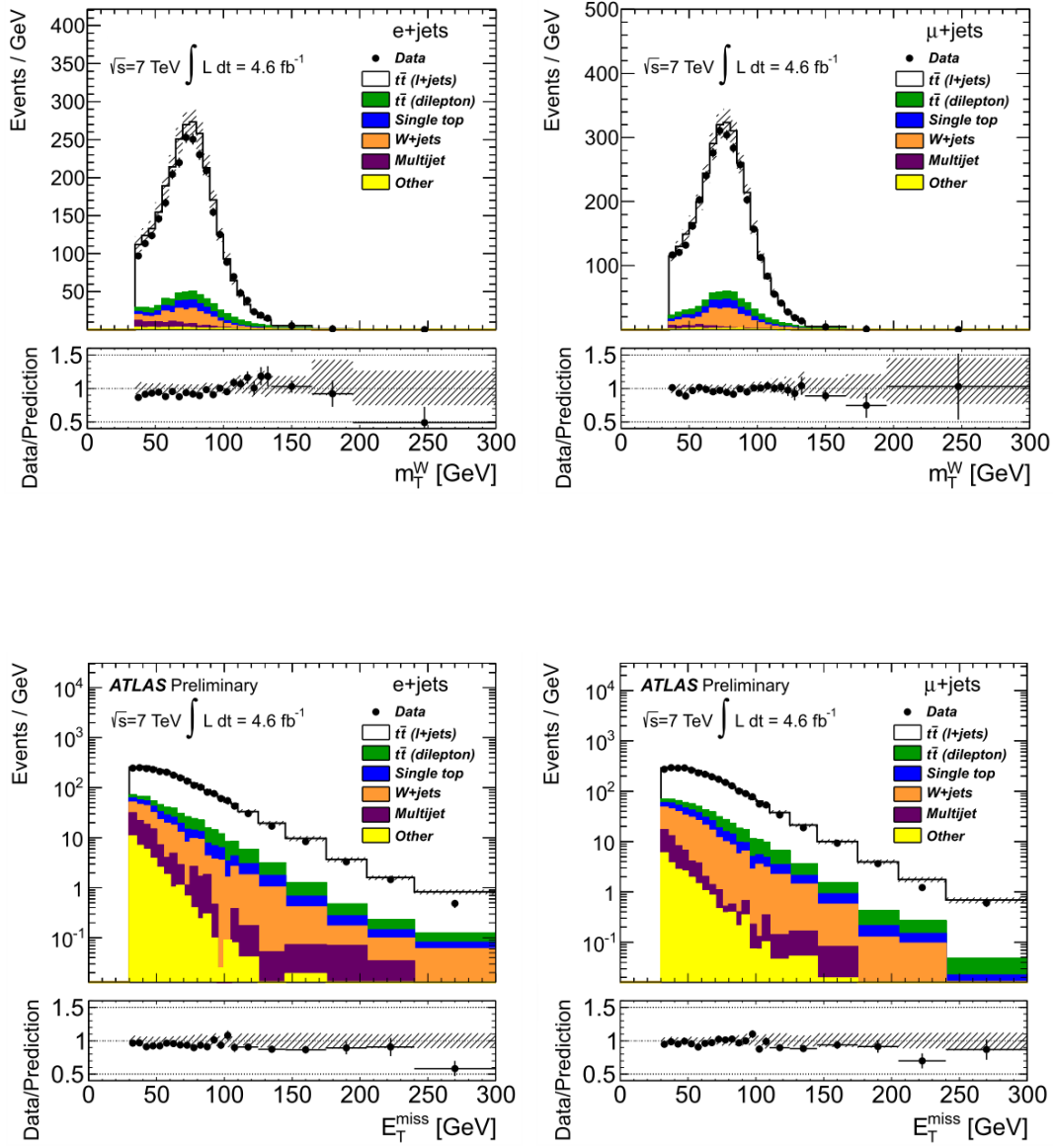


Figure 5.3 – Reconstructed distribution for m_T^W in the electron (upper-left) and muon (upper-right) channels and for E_T^{miss} in the electron (bottom-left) and muon (bottom-right) channels. Data is compared to predictions, using Alpgen + Herwig as the signal model. The hashed area indicates the combined statistical and systematic uncertainties on the total prediction, excluding systematic uncertainties related to the modeling of the t \bar{t} system. “Other” includes the small backgrounds from the diboson and Z + jets production.

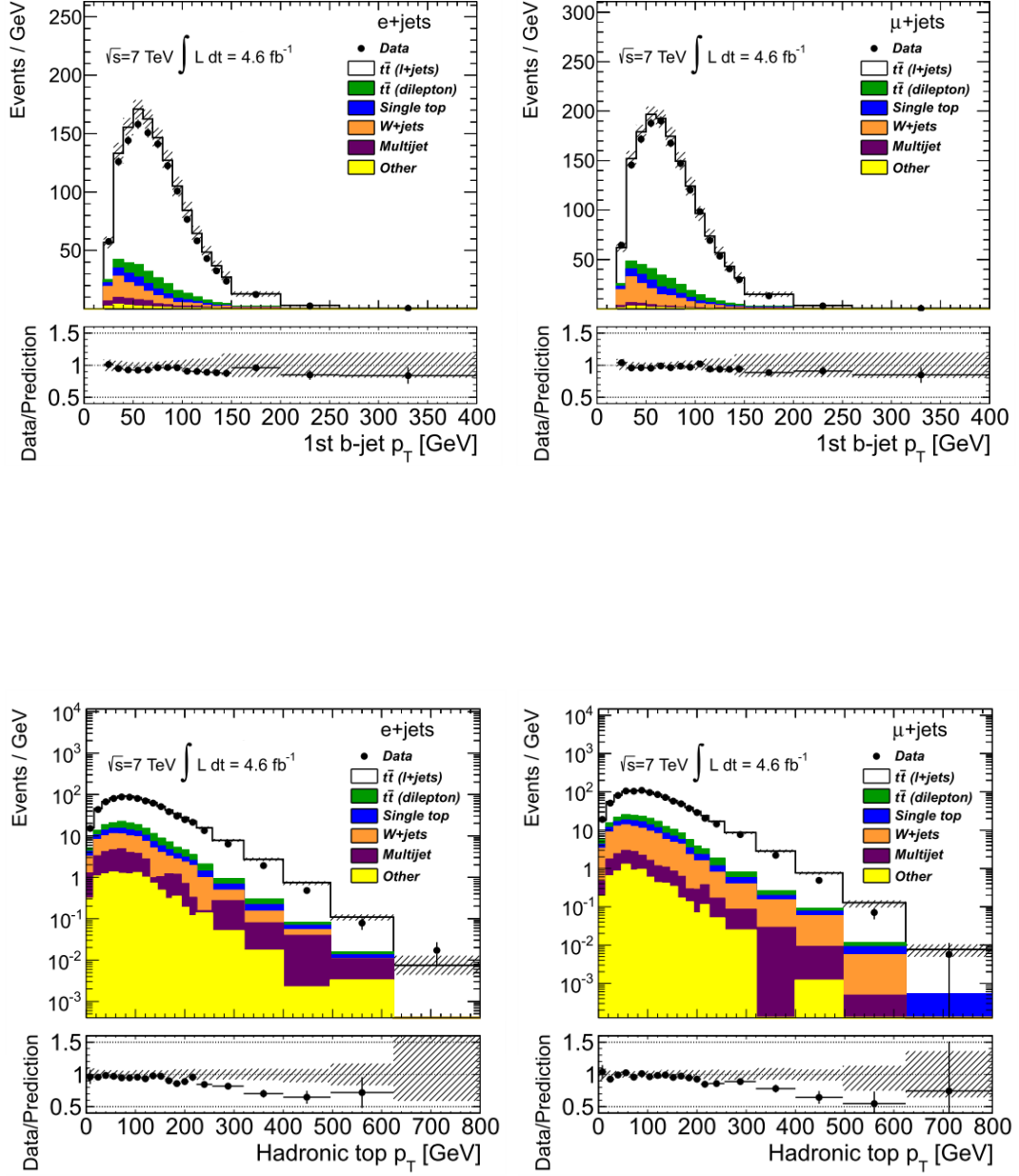


Figure 5.4 – Reconstructed distribution for p_T of the leading- p_T b -jet in the electron (upper-left) and muon (upper-right) channels and for hadronic top p_T in the electron (bottom-left) and muon (bottom-right) channels. Data is compared to predictions, using Alpgen + Herwig as the signal model. The hashed area indicates the combined statistical and systematic uncertainties on the total prediction, excluding systematic uncertainties related to the modeling of the $t\bar{t}$ system. "Other" includes the small backgrounds from the diboson and Z + jets production.

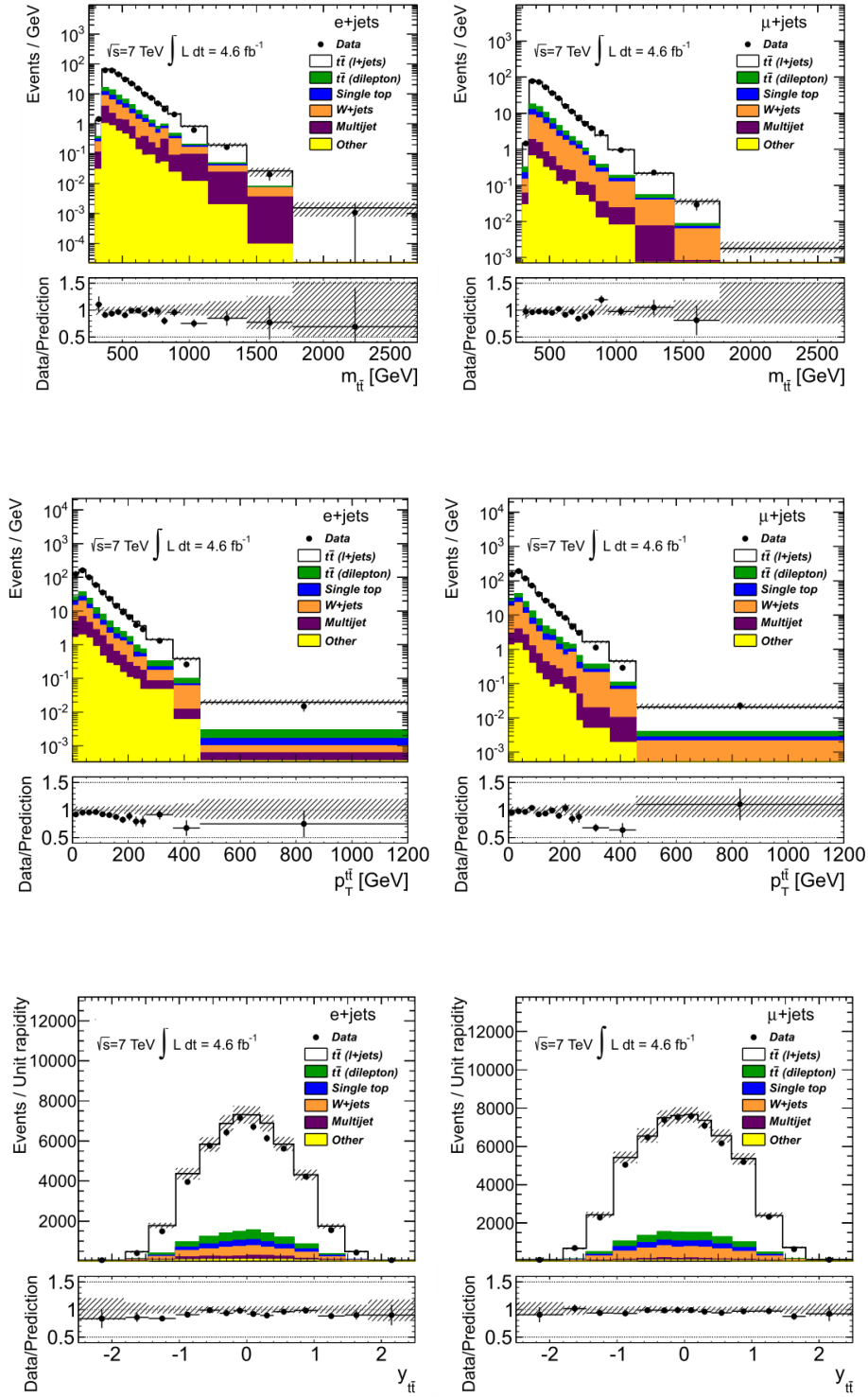


Figure 5.5 – Reconstructed distribution for $m_{t\bar{t}}$ in the electron (upper-left) and muon (upper-right) channels, for $p_T^{t\bar{t}}$ in the electron (center-left) and muon (center-right) channels and for $y_{t\bar{t}}$ in the electron (bottom-left) and muon (bottom-right) channels. Data is compared to predictions, using Alpgen+Herwig as the signal model. The hashed area indicates the combined statistical and systematic uncertainties on the total prediction, excluding systematic uncertainties related to the modeling of the $t\bar{t}$ system. “Other” includes the small backgrounds from the diboson and $Z + \text{jets}$ production.

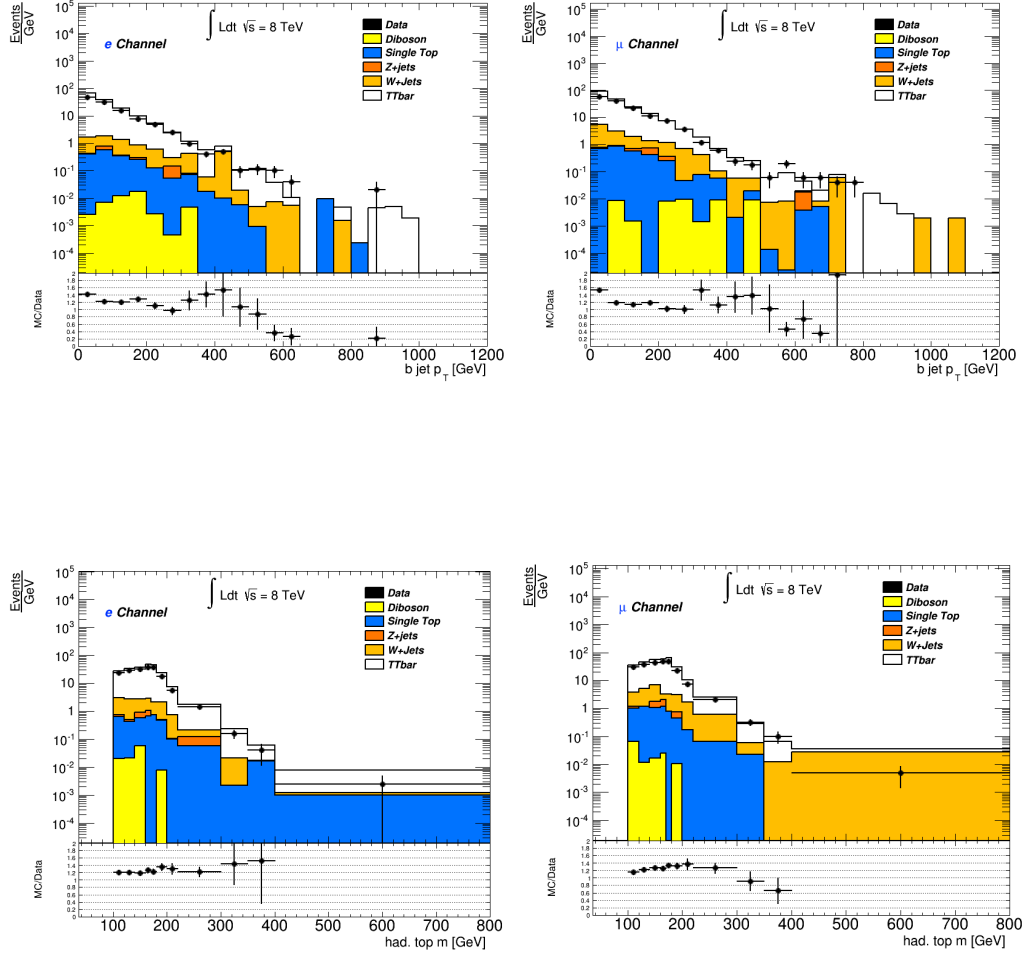


Figure 5.6 – Reconstructed distribution for $m_{t\bar{t}}$ in the electron (upper-left) and muon (upper-right) channels and for $p_T^{t\bar{t}}$ in the electron (bottom-left) and muon (bottom-right) channels. Data is compared to predictions, using Alpgen+Herwig as the signal model. The hashed area indicates the combined statistical and systematic uncertainties on the total prediction, excluding systematic uncertainties related to the modeling of the $t\bar{t}$ system. “Other” includes the small backgrounds from the diboson and $Z + \text{jets}$ production.

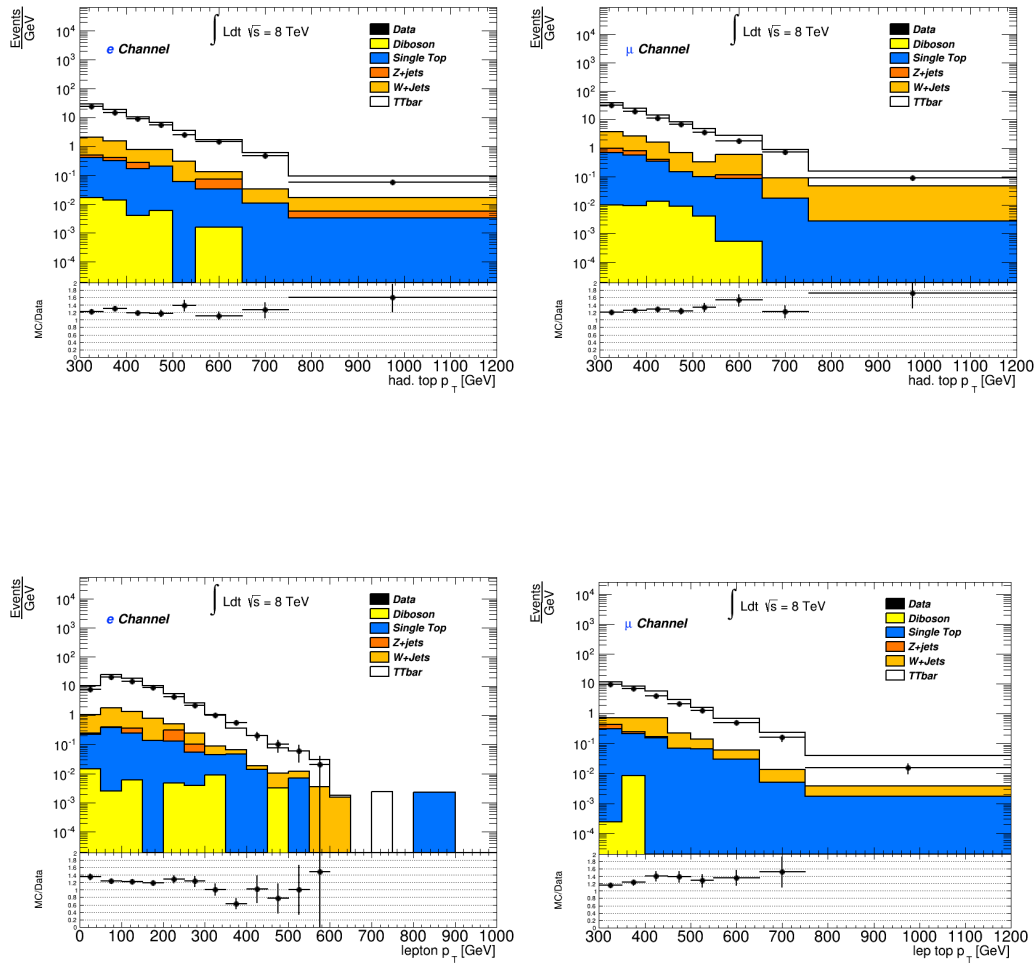


Figure 5.7 – Reconstructed distribution for the hadronic top p_T in the electron (upper-left) and muon (upper-right) channels and for the leptonic top p_T in the electron (bottom-left) and muon (bottom-right) channels. Data is compared to predictions, using Alpgen+Herwig as the signal model. The hashed area indicates the combined statistical and systematic uncertainties on the total prediction, excluding systematic uncertainties related to the modeling of the $t\bar{t}$ system. “Other” includes the small backgrounds from the diboson and Z + jets production.

5.4 Unfolding

In high-energy physics, measurements of physical observables are usually distorted or biased by the limited resolution and acceptance of the detector; this prevents a direct comparison of the measured distributions with the theoretical predictions. In order to make that possible, the data must be corrected for these detector effects via a procedure called *unfolding*. The acceptance of the detector is defined as the probability to observe an event in a certain kinematic region. It is quite easy to deal with this effect by applying a bin dependent scale factor, evaluated from Monte Carlo, to rescale data to the predicted number of events. Instead, the finite resolution effect is due to the limited accuracy of the detector and leads to a statistical smearing between the true kinematical variable x and the measured quantity y . This effect can be mathematically represented by the *Fredholm integral*

$$g(y) = \int A(y, x) f(x) dx \quad 5.3$$

defining $g(y)$ the measured distribution and $f(x)$ the true one. The resolution function $A(y, x)$ represents the distortion introduced by the detector. The unfolding consists in evaluating the true distribution $f(x)$ from a given $g(y)$ and $A(y, x)$; this is the so called *inverse problem* that requires to calculate the inverse of the resolution function $A^{-1}(y, x)$ to be solved. Moreover, in high energy physics applications, the above approach is usually applied to discrete distributions so the resolution function becomes a discrete matrix evaluated by Monte Carlo simulation, called *Response Matrix* (or *Migration Matrix*). In this case, integral 5.3 is replaced by a system of linear equations;

the inversion problem is then aggravated by statistical and possible systematic errors of the *response matrix* itself.

The inversion of a finite system of equation rarely admits an exact solution, so a manifold of different techniques calculating approximate solutions have been developed. In order to overcome the intrinsic instability of these approximate solutions, some kind of regularization conditions can be imposed, based on some *a priori* information as the request of minimum curvature or that the solution must be strictly positive [59]. These methods usually allows the suppression of spurious oscillating components of the unfolded solution and often lead to satisfactory results, even if practical implementations may become quite heavy and labyrinthic.

5.4.1 Unfolding Methods

Four unfolding methods have been taken into account in these analyses, being the most diffuses in high energy physics:

BIN-BY-BIN procedure extracts correction factors for each bin from the ratio of the reconstructed simulation over the theoretical predictions distributions. The bin-by-bin is the simplest method and cannot be even strictly considered a proper unfolding technique because it does not consider migration effects. Anyway, if the response matrix is close to diagonal, it provides a good approximation of the true value and is then useful as a cross check for more complex techniques.

SIMPLE MATRIX INVERSION approaches the unfolding problem by its very basic definition. Considering x_j the vector of true generated events in each bin j and y_i the vector of measured data in each bin i , they are related by the *migration matrix* \mathcal{M}_{ij} as follows:

$$y_i = \mathcal{M}_{ij} \cdot x_j \quad 5.4$$

By numerically inverting the matrix the measured data can be unfolded to the true values.

ITERATIVE BAYESIAN can be viewed as a “cause and effect” procedure [60].

Causes C_i correspond to the generated true values while effects E_j are the events effectively measured by the apparatus. Each cause can produce different effects, but for a given effect the exact corresponding cause is not known. However, the probability $P(E_j | C_i)$ for a cause to generate a specific effect can be evaluated assuming some *a priori* knowledge, usually accomplished by Monte Carlo simulations. The Bayes Theorem allows estimating the probability $P(C_i | E_j)$ that cause C_i is the real responsible for the effect E_j

$$P(C_i | E_j) = \frac{P(E_j | C_i) \cdot P_0(C_i)}{\sum_{k=1}^{n_C} P(E_j | C_k) \cdot P_0(C_k)} \quad 5.5$$

where $P_0(C_i)$ is the prior probability for the cause C_i and n_C is the number of possible causes. The estimator for the number of causes in the $i - th$ bin can be expressed as

$$\hat{n}(C_i) = \frac{1}{\varepsilon_i} \sum_{j=1}^{n_E} n(E_j) \cdot P(C_i | E_j) \quad 5.6$$

where $n(E_j)$ is the number of measured effects in the $j - th$ bin and ε_i is the efficiencies for each bin, evaluated from Monte Carlo. These efficiencies are defined as

$$\varepsilon_i = \sum_{j=1}^{n_E} n(E_j)^{MC} / n(C_i)^{MC} \quad 5.7$$

The $P(C_i | E_j)$ can be considered equivalents to the elements of an inverted migration matrix

$$\mathcal{M}_{ij} = \frac{P(E_j | C_i) \cdot P_0(C_i)}{\sum_{h=1}^{n_E} P(C_i | E_h) \cdot \sum_{k=1}^{n_C} P(E_j | C_k) \cdot P_0(C_k)} \quad 5.8$$

leading to the following unfolding equation

$$\hat{n}(C_i) = \sum_{j=1}^{n_E} \mathcal{M}_{ij} n(E_j) \quad 5.9$$

The new causes' probability

$$P'_0(C_i) = \frac{\hat{n}(C_i)}{\sum_{i=1}^{n_C} \hat{n}(C_i)} \quad 5.10$$

replaces the initial prior $P_0(C_i)$ and the procedure is reiterated until $P_0(C_i)$ and $P'_0(C_i)$ converge.

SINGLE VALUE DECOMPOSITION (SVD), or Tichonov regularization, [61]

is an extension of the simple inversion matrix that uses a regularization technique in order to reduce possible rapidly oscillating solutions due to numerical calculation instabilities and finite samples statistics. The n -dimensions migration matrix can be written as

$$\mathcal{M} = \mathcal{U}\mathcal{S}\mathcal{V}^T \quad 5.11$$

where \mathcal{U} is an $n \times n$ orthogonal matrix, \mathcal{V} is an $n \times n$ orthogonal matrix and \mathcal{S} is an $n \times n$ diagonal matrix. Then the inverted migration matrix can then be evaluated as

$$\mathcal{M}^{-1} = \mathcal{U}\mathcal{S}^{-1}\mathcal{V}^T \quad 5.12$$

that greatly simplifies the inversion process. The exact solution of the unfolding system in Equation 5.4 leads in most of the cases to a rapidly oscillating distribution. The oscillatory component coming out from the calculation can be suppressed using some *a priori* knowledge about the solution. Technically this can be achieved by adding a regularization term to the system and minimizing it

$$(\mathcal{M}x - y)^T(\mathcal{M}x - y) + \tau \cdot (\mathcal{C}x)^T \mathcal{C}x = \min. \quad 5.13$$

Here τ is the regularization parameter that determines the relative weight of the *a priori* condition on the solution defined by the matrix \mathcal{C} . In particle physics the \mathcal{C} matrix is usually chosen from the theoretical distributions obtained from MC simulations, while the optimal value of τ is problem-

dependent and must be determined from case to case. In order to reduce the oscillation of the solution, the x distribution should be smooth, with small bin-to-bin variations. The solution of this new equation system can be calculated using the SVD inversion procedure that rotates vectors and matrices in a different phase space by the transformations

$$\mathcal{U}\mathcal{S}\mathcal{V}^T = M\mathcal{C}^{-1}, \quad d = \mathcal{U}^T y, \quad z = \mathcal{V}^T \mathcal{C}x \quad 5.14$$

The Fourier expansion of the vector d is used in the calculation; for reasonably smooth initial distribution y , only the first few d_i terms are supposed to be significant then it is convenient to trunk the expansion in order to avoid a rapidly oscillating solution. The k factor is the number of terms that are kept unsuppressed and it is an input variable of the method. A small value of the k factor means a strong regularization, losing information, and giving an higher weight to the MC condition \mathcal{C} , while a large value of k gives less importance to the *a priori* knowledge on the distribution, but allows more oscillating terms in the solutions. The k factor is also related to the regularization parameter τ that define “how strong” should be the impact of the *a priori* condition settled up in the regularization procedure; the regularization parameter τ and the k factor are bounded by the equation

$$\tau = \left[\frac{d_k}{z_k} \right]^2 \quad 5.15$$

where k is the k factor and the vectors d_k and z_k are defined in Equation 5.14. The k factor value should be decided depending on the case under study by

looking at the d_i terms distribution. A good k factor can be chose equal to the first value of i for which the expansion term d_i starts to become negligible.

5.4.2 Unfolding Studies

The four methods described above have been compared one to each other in order to select one of them to be used in the analysis. The comparison has been accomplished by looking at the difference between the unfolded differential cross section spectra for reconstructed and generated signal distributions both coming from Monte Carlo simulations. This is done with the formula

$$diff_x = \frac{\left[\frac{d\sigma}{dx} \right]_{reco} - \left[\frac{d\sigma}{dx} \right]_{gen}}{\left[\frac{d\sigma}{dx} \right]_{gen}} \quad 5.16$$

where x is the considered kinematic variable on which the cross section distribution runs.

The generated sample used to “train” the unfolding algorithm is formed by half of the $t\bar{t}$ Monte Carlo sample. The reconstructed events on which the unfolding is applied to, are derived from the other half of the MC sample where the numbers of events in each bin of the reconstructed distributions has been randomly smeared following a Gaussian distribution before the unfolding process. The unfolding has been performed on 5000 different smeared “toys” distributions and the final spread on the averaged result forms the statistical error. As can be inferred by plots in Figure 5.8 obtained from

the 2011 data, all methods perform reasonable well, but the SVD procedure results slightly more stable in the tail of the distribution; for that reason it has been chosen to be used in the final measurement. The test has been performed using three iterations for the Bayesian unfolding. The choice of the regularization parameter k in the SVD method is done in a way that the measurement remains unbiased and the minimum amount of information is lost, but however preventing rapidly oscillating solutions.

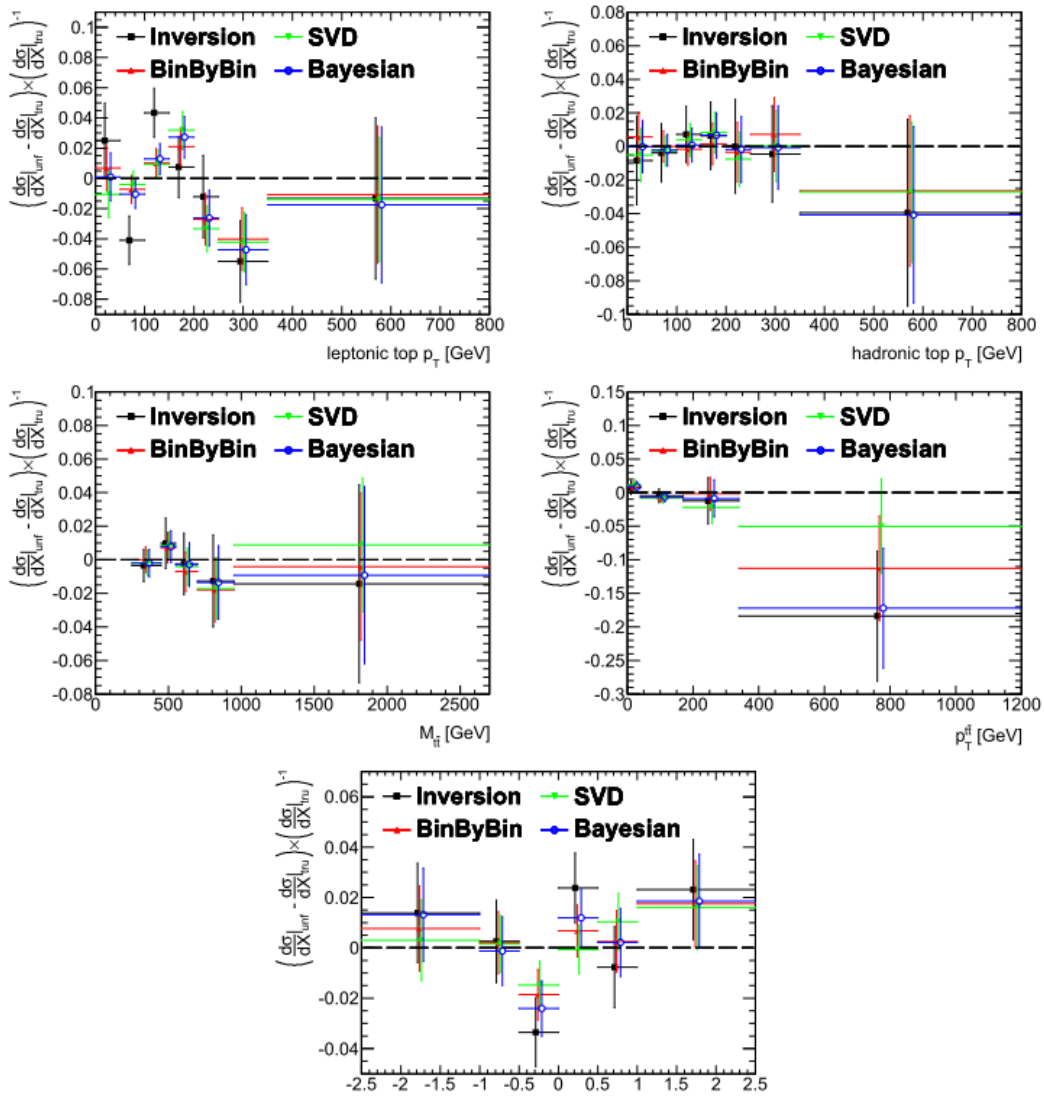


Figure 5.8 – Comparison of the four unfolding methods for each variable using two statistically independent halves of the MC@NLO sample in the $e+jets$ channel.

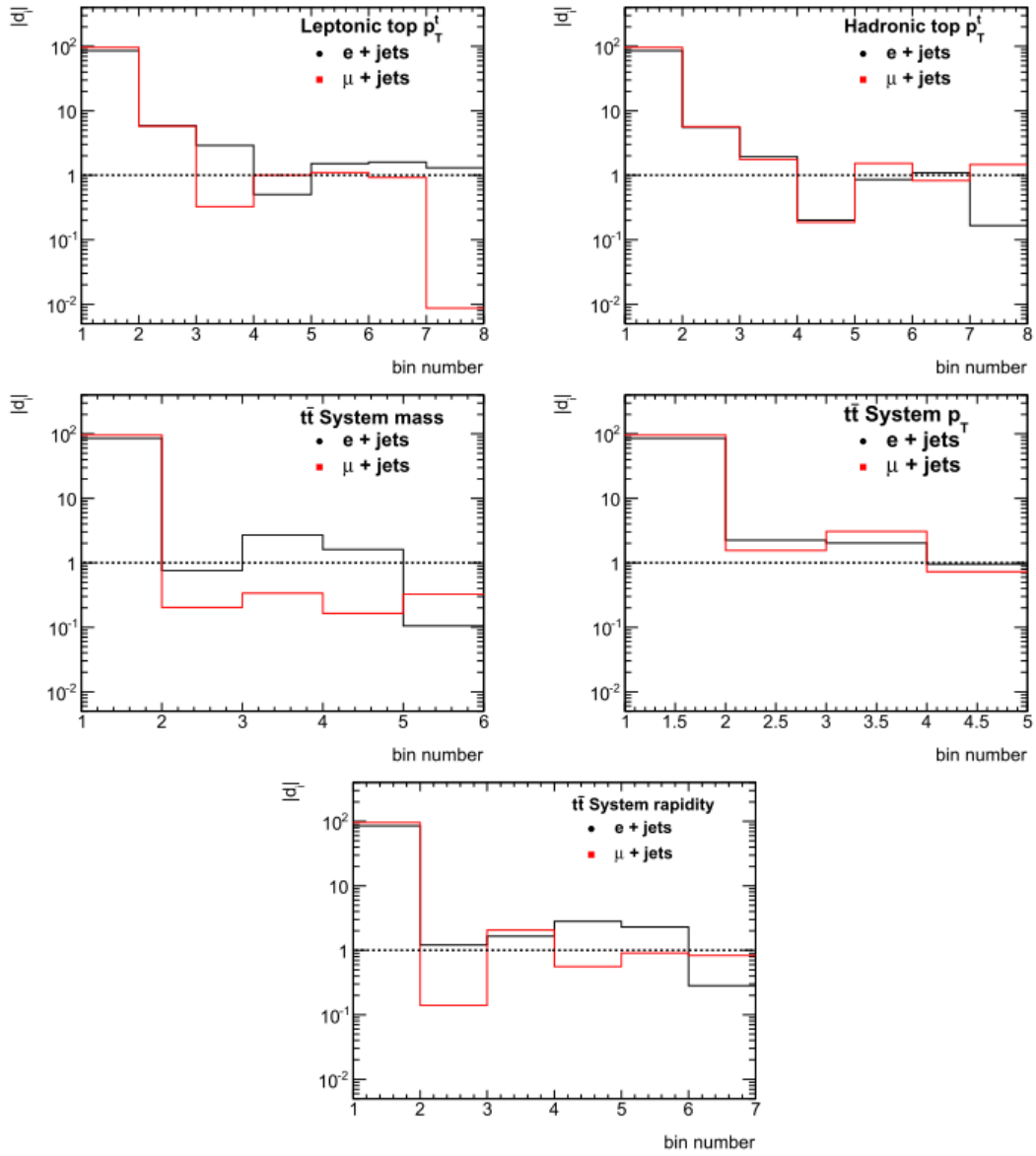


Figure 5.9 – The “ $|d_i|$ ” plot, i.e. the distribution of the transformed measured vector in the diagonalized space, ordered in descending eigenvalues in the SVD method, as function of bin number in the measured distribution. This distribution is used to select the correct regularization parameter to use in the SVD unfolding method. Signal has been modeled by ALPGEN.

In order to decide the best value for the k factor, the distributions $|d_i|$ of the decomposed measured events in the SVD diagonalized space, ordered in descending eigenvalues λ_i , should be considered [61] (see Figure 5.9). The best k factor value is equal the smaller bin that guarantee a sufficiently small $|d_i|$ oscillation.

In the 2011 data analysis the value of $k = 4$ has been chosen for the $y^{t\bar{t}}$ and the top p_T^t spectrums, while $k = 3$ is fixed in the case of $p_T^{t\bar{t}}$ and $m^{t\bar{t}}$ spectrums. In the 2012 data analysis the regularization factor $k = 5$ has been adopted for the hadronic top p_T^t distribution; the decision is taken looking also to the statistical and systematical uncertainties because of the not conclusive result from the $|d_i|$ distributions.

5.4.3 Parton and Particle Level

The theoretical predictions and the unfolded results can be compared at two different levels: the *parton* and *particle* levels, as already mentioned in Chapter 1. In the *parton level* case the generated top partons are compared with the measured and unfolded distributions. The measurement is performed in the full phase-space, using the MC to extrapolate results from the region visible by the detector to the full physical phase-space. In the *particle-level* case, the unfolded objects are compared to similar objects defined at the truth level using stable particles produced after hadronization. The comparison between unfolded events and MC truth ones, is done in a limited phase-space region which definition should match as much as possible with the one accessible by the detector. This is accomplished by applying to the truth partners of the objects

used in the offline analysis, a selection similar to the one applied on reconstructed events. In the case of unfolding at *particle level*, only detector effects are expected to play a role, therefore the *migration matrix* results close to the diagonal. In the case of a *parton level* unfolding it is expected a changing in the *migration matrix* for different Monte Carlo generators, because also different generation models may play a role. The truth objects used in the *particle level* selection are defined as follows:

LEPTONS must not to come from a quark or an hadron. The leptons are dressed

by summing up all stable photons within a cone of radius $R = 0.1$.

JETS are reconstructed using the anti- k_t algorithm, starting from all stable truth particles with the exception of those indirectly matched to a W boson.

MISSING TRANSVERSE ENERGY is given by the sum of all neutrinos coming from a W boson.

In the 2011 data analysis the unfolding is done at the *parton level*, while in the 2012 data analysis the collected events are unfolded at the *particle level*, by the definition of a *fiducial phase-space* via the following requirements:

- Exactly one *dressed* lepton with $p_T > 25 \text{ GeV}$ and $|\eta| < 2.5$. The electron (muon) should be far enough from all *small-R* jets, $\Delta R_{lepton,jet} > 0.4$ (0.1);
- The Overlap Removal procedure for muons, *small-R jets* and electrons(see Paragraph ...);
- At least one *small-R* jet with $p_T > 25 \text{ GeV}$ and $|\eta| < 2.5$ within $\Delta R < 1.5$ from the lepton;

- At least one trimmed *large-R* jet with $p_T > 300 \text{ GeV}$, $m > 100 \text{ GeV}$, $\sqrt{d_{12}} > 40 \text{ GeV}$ and $|\eta| < 2$. The additional requirements to be well separated from both the lepton ($\Delta\phi > 2.3$) and the *small-R* jets associated with the leptonic top ($\Delta R > 1.5$);
- $E_T^{\text{miss}} > 30 \text{ GeV}$ and $m_T^W > 30 \text{ GeV}$ for the electron channel;
- $E_T^{\text{miss}} > 20 \text{ GeV}$ and $E_T^{\text{miss}} + m_T^W > 60 \text{ GeV}$ for the muon channel;
- At least one *b-tagged small-R* jet with $p_T > 25 \text{ GeV}$ and $|\eta| < 2.5$.

A negligible, less than 5%, dilepton event contamination survives to the *particle-level* selection as defined here.

5.4.4 Migration Matrices

The migration matrixes evaluated from Monte Carlo and used in the unfolding procedure are reported below for each studied variable, from Figure 5.10 to Figure 5.14. All the matrices are close to the diagonal that indicates a relatively small migration effect; this is particularly true in the 2012 data case (Figure 5.14) thanks to the *particle level* definition. All matrices are normalized with respect to the number of truth events.

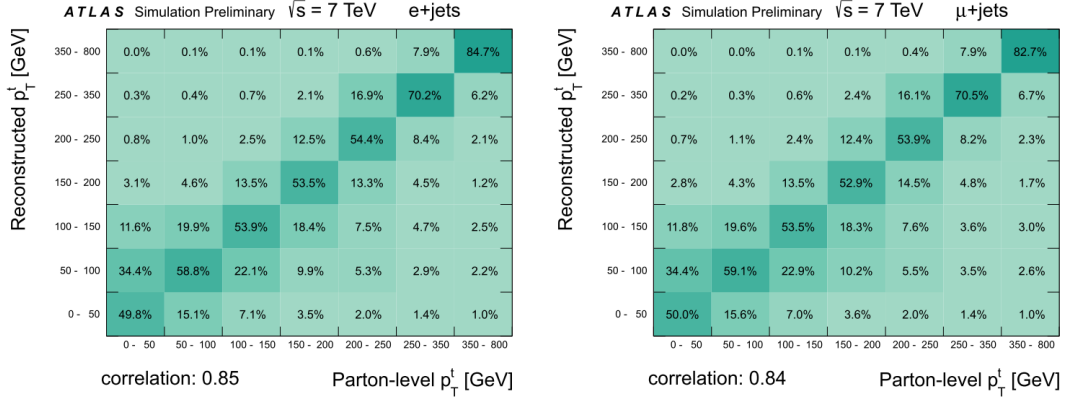


Figure 5.10 – The migration matrices obtained from the Alpgen + Herwig simulation in the 2011 analysis, relating the parton and reconstruction levels for p_T^l in the electron (left) and muon (right) channels. The linear correlation coefficient is given below each plot and all columns are normalized to unity.

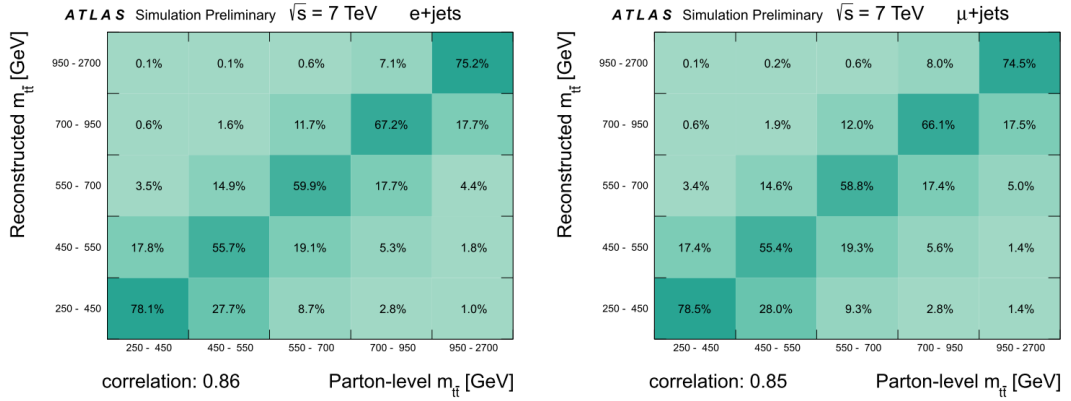


Figure 5.11 – The migration matrices obtained from the Alpgen + Herwig simulation in the 2011 analysis, relating the parton and reconstruction levels for $m_{t\bar{t}}$ in the electron (left) and muon (right) channels. The linear correlation coefficient is given below each plot and all columns are normalized to unity.

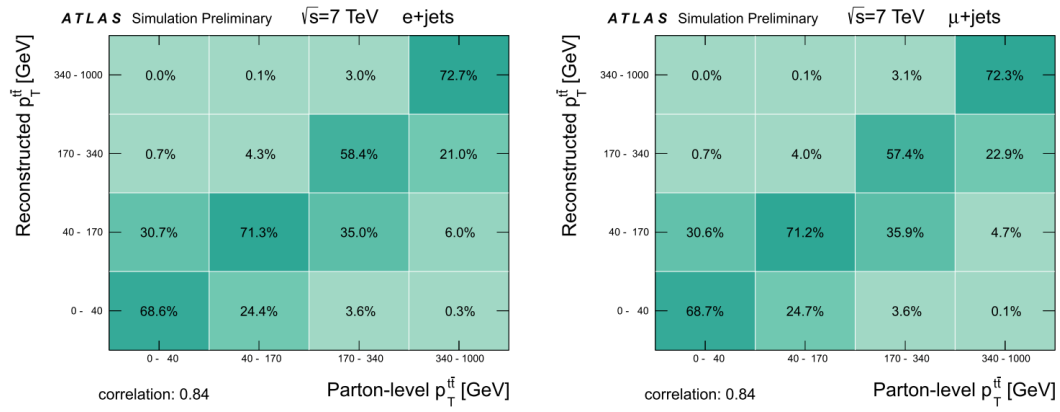


Figure 5.12 – The migration matrices obtained from the Alpgen + Herwig simulation in the 2011 analysis, relating the parton and reconstruction levels for p_T^t in the electron (left) and muon (right) channels. The linear correlation coefficient is given below each plot and all columns are normalized to unity.

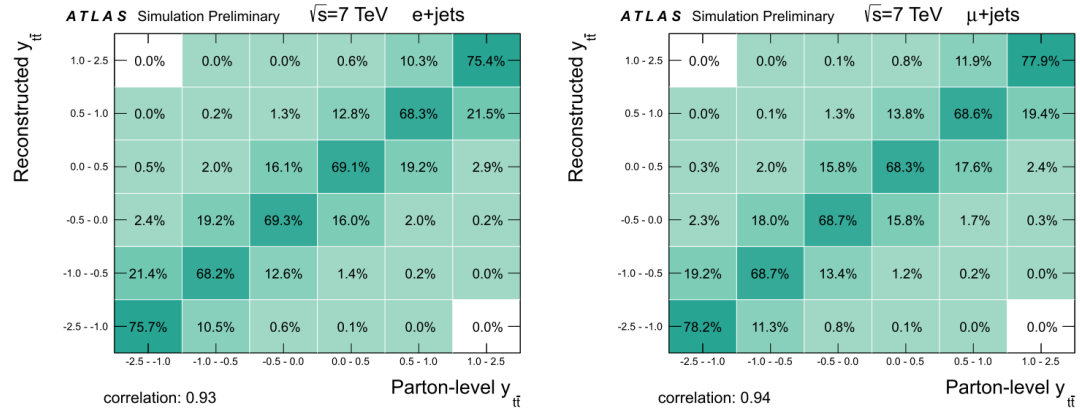


Figure 5.13 – The migration matrices obtained from the Alpgen + Herwig simulation in the 2011 analysis, relating the parton and reconstruction levels for y_t in the electron (left) and muon (right) channels. The linear correlation coefficient is given below each plot and all columns are normalized to unity.

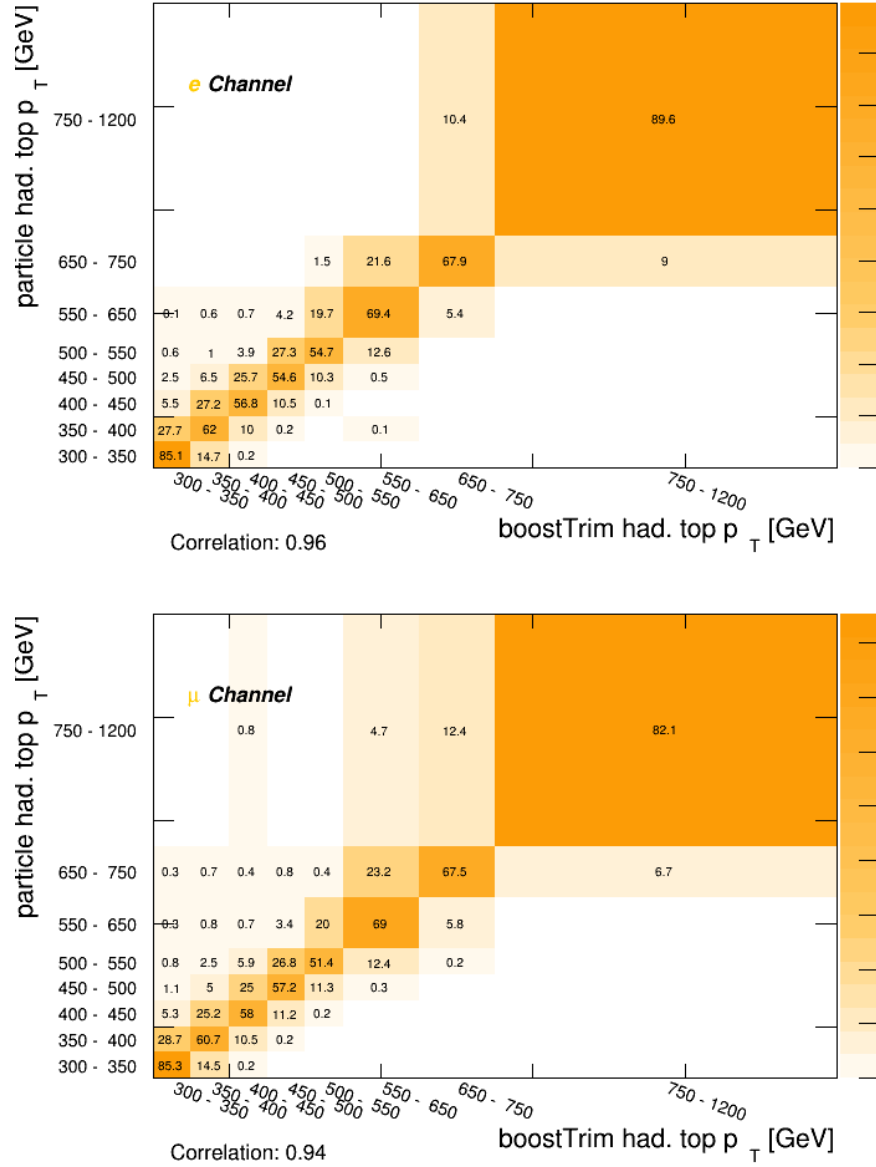


Figure 5.14 – The migration matrices obtained from the Pythia+Herwig simulation in the 2012 analysis, relating the particle and reconstruction levels for p_T^{τ} in the electron (up) and muon (bottom) channels. The linear correlation coefficient is given below each plot and all columns are normalized to unity.

5.4.5 Unfolding Error Propagation

The effects of systematics uncertainties on the unfolded distributions have been evaluated by applying artificial variations to the correspondent MC distributions (i.e. background, signal efficiency and migration matrix) that are used to unfold reconstructed data. The uncertainty on the measurements is then calculated as the difference between the *nominal* unfolding result and the unfolded distribution with systematic variations. The total systematic uncertainty in each bin is calculated by summing all deviations in quadrature, depending on its sign (positive deviations contribute to the “up” uncertainty while negative to the “down” uncertainty). In the 2011 data analysis the *initial and final state radiation* (I/FSR) systematic follows a different estimation prescription, since the corresponding MC samples have been generated with a LO generator, not compatible with the NLO reference. In this case, the systematic uncertainties is considered as the relative difference between the two unfolded spectra coming from the MC samples, respectively perturbed in the one sense and in the other. A detailed summary of the uncertainty coming from each systematic source considered is collected in Chapter 6 and Appendix A.

5.4.6 Channel combination

In the 2011 analysis the combination of the electron and muon channels has been performed in order to improve the sensitivity of the final result. The measured cross section distributions evaluated by the two channels are combined with the Best Linear Unbiased Estimator (BLUE) method; it consists in a weighted mean that includes the full covariance matrix between the channels. It is constructed by assuming zero or full correlation for channel-specific or common systematic uncertainty sources, respectively. The method allows to take into account only systematic uncertainties with symmetric contribution so the average of the positive and negative error is used.

The combined results have been compared and a good agreement has been found with respect to the results obtained from the two channels separately. The combination result has been also compared with the one obtained from the simple merging of the two channels. Neither the results from individual electron and muon channels consistently differ each other: the differences observed in the corresponding bins for all variables of interest are within two standard deviations.

5.5 Analysis framework

In a huge collaboration as ATLAS the transfer knowledge is a key point for an efficient teamwork in order to share tools for common tasks. This is possible only through a code with a friendly user interface and a high compatibility and flexibility.

On this prospective I have designed the analysis infrastructure used for the 2012 analysis called BoostedRealm: it is fully integrated in the most common ATLAS top

analysis infrastructure named TopRootCore (TRC) that is derived and inherits from the most general RootCore ATLAS framework. The inclusion in TRC guarantees to easy maintain corrections and scale factors up to date and gives the possibility to use several tools exploiting for different analysis tasks already implemented by other groups.

The BoostedRealm framework has a modular structure trough a strong hierarchical coding. The different analysis steps, in order to simplify the managing and to be as much user friendly as possible, are managed via bash scripts and parameter files. A TRC package called TopD3PDBoosted, of which I am one of the developers, is included in the framework and deals with the events correction, selection and reconstruction; it is central validated from the ATLAS top working-group. The package is currently used in our $t\bar{t}$ differential cross section measurement as well as in the $t\bar{t}$ resonances search. The BoostedRealm framework also includes some other TRC independent programs dealing with the main analysis tasks as unfolding and systematic treatments.

Specifically, the analysis program covers all the phases of the data analysis: it starts processing data in D3PD format, both on the GRID or using local batch systems, by applying the event selection and by producing slim and easy manageable ntuple files. This phase is accounted by the TopD3PDBoosted package. I have added to the same package a program dealing with the second analysis step that processes the ntuple for further selection requests and then reconstructs the physical objects producing row histograms. I have included in the TopD3PDBoosted package the TOM algorithm (see Chapter 4) in order to allows its spread among the ATLAS collaboration; in fact several groups have already shown great interest to use it in their analysis. An independent code is given in order to produce final plots in the ATLAS format, as for data/MC comparison. A flexible unfolding tool is also included; it gives the possibility to use different

unfolding techniques and to produce cross section measurements; a flexible system to manage statistical and systematic uncertainties is also included. A series of scripts and macros deal with data manage and storage. The framework structured in a way that allows to be easily adapted to different top analysis to add new features. The framework is equipped with a documentation describing the different programs included and guiding the analyzer to perform a full basic analysis chain.

Chapter 6

Results

This chapter will present the results about the differential production cross section of $t\bar{t}$ pairs obtained from the analysis of the 2011 and 2012 data samples collected with the ATLAS detector. The techniques used in order to estimate the statistical and systematic uncertainties affecting these measurements are described in Paragraphs 6.1 and 6.2; final results are shown in Paragraphs 6.3 and 6.4.

6.1 Statistic uncertainty

The statistical uncertainty on the differential cross section measurements is due to the finite number of data events collected. It is evaluated by repeating the same analysis procedure, comprehensive of unfolding, by the using of *pseudo-experiments*, which are simulated distributions obtained by smearing the bin-per-bin value of the background-subtracted data distribution according to a Poissonian distribution. The uncertainty on the final distribution in each bin is calculated as the root mean square of the observed cross section distributions for all pseudo-experiments. It will be summed in quadrature with the total systematic uncertainty.

6.2 Systematic uncertainty

Systematic uncertainties are originated by both an imperfect knowledge of the detector, that entails an uncertainty on the parameters used in the event reconstruction, and in the theoretical modeling of signal and background events. Each systematic uncertainty is evaluated by varying the nominal distribution by one standard deviation on the overall effect caused by the considered error source; this usually leads to two shifted distributions representing the fluctuation with respect to the nominal distribution. The bin-per-bin uncertainty has been evaluated as the difference between the nominal differential cross section and the shifted one, both calculated after the event selection

and the unfolding process. In cases where both shifted distributions produce an excess or a defect with respect to the nominal distribution, the resulting variation is assumed to be of the same size in both directions and is therefore symmetrized.

While in the 2011 analysis all the systematic uncertainties affecting the $t\bar{t}$ signal have been already considered, in the more recent 2012 analysis only a limited set of systematic sources have been estimated. In particular, two sources which contribution strength have been evaluated as one of the most significant from similar-selection analysis, still missing. For this reason, in this thesis, the main result will be presented without including the systematic uncertainty contribution that will be instead shown in Appendix A.

The various systematic uncertainty sources affecting the 2011 and 2012 analyses have been divided in different categories depending on which parts of the simulation are involved: the signal, the background or the detector modeling.

6.2.1 Signal modeling

The sources of uncertainty affecting the $t\bar{t}$ signal modeling described in this section are: the hard process simulated in the chosen MC generator, the parton shower, the final and initial state radiation (IFSR) setting and the choice of the PDF set.

MONTE CARLO GENERATOR

The uncertainties in the Monte Carlo hard process simulation are evaluated from the difference between final distributions produced by different

generators. In the 2011 analysis the *ALPGEN* generator is used for the nominal sample and *MC@NLO* is considered for the evaluation of the MC systematic uncertainty; in both cases Herwig is used to simulate the parton showering. In the 2012 analysis the PowHeg+Pythia generator is used for the nominal sample while *MC@NLO*+Herwig is considered for comparison.

PARTON SHOWER

The systematic contribution from the parton shower modeling uncertainty is evaluated by comparing results from *ALPGEN*+Herwig/Jimmy and *ALPGEN*+Pythia samples. This contribution is not yet considered in the 2012 analysis.

INITIAL AND FINAL STATE RADIATION

In the 2011 analysis the initial and final state radiation (IFSR) uncertainty is evaluated with dedicated AFII-*ALPGEN*+Pythia MC samples obtained varying radiation settings consistently with the α_s variation both in the hard matrix element and in the parton shower, using the dedicated Perugia 2011 tune [78] of Pythia. The shifted samples are then unfolded using the *ALPGEN*+Herwig sample and the uncertainty is evaluated by taking half the difference with respect to the nominal distributions. In the 2012 analysis the same Powheg+Pythia generator is used both to estimate the effect of the radiation

setting change and to produce the nominal samples used in the unfolding process.

PDF

The uncertainty related to the choice of a particular parton distribution function during the event simulation has been studied by reweighting the events with different sets of NLO PDF in order to see the effect on the event distribution. Three sets of PDF has been considered: MSTW08, CT10 and NNPDF2.0. The difference between the maximum and the minimum value of the resulting three distributions, comprehensive of one standard deviation σ error variation with respect to the central value, is called envelope; the uncertainty value is considered as the half of the maximum envelope extension. In Figure 6.1 the envelope of the cross section multiplied for the generated luminosity ($\sigma \times \mathcal{L}$) evaluated in the 2011 analysis is shown; studies to determine the PDF uncertainty in the 2012 analysis are still ongoing using the same method.

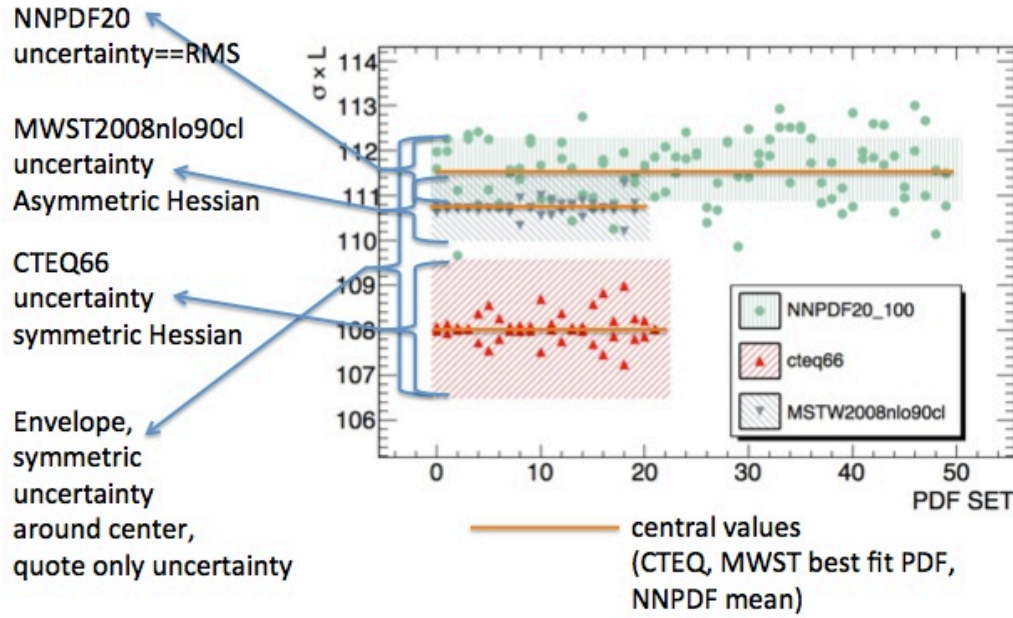


Figure 6.1 – Definition of the PDF uncertainty envelope (data from the “dilepton” analysis in the ee channel).

6.2.2 Background modeling

The description of the uncertainty related to the main data-driven backgrounds (W +jets and QCD) is given. The contributions to the overall systematic uncertainty by other backgrounds have been estimated to be negligible because of their small effect.

W +JETS

The W +jets contribution to systematic uncertainties is evaluated varying both the data-driven normalization and the heavy flavour factors within their uncertainties that are given by the data-driven method itself. These data-

driven uncertainties on the applied scale factors are shown in Table 6.1 for the 2011 and 2012 data respectively.

QCD

The QCD background uncertainty is estimated by comparing data and multi-jets *data-driven* background in different control regions. The resulting normalization uncertainty has been estimated to be 50% in the $e + jets$ channel and 20% in the $\mu + jets$ one. In the 2012 analysis, the studies on the evaluation of the data driven background as well as its uncertainties from specific control regions is still ongoing and will not be presented here.

Channel	F_{bb}, F_{cc}	F_c	F_{ll}	Overall normalization factor
$e+jets$	1.41 ± 0.35	0.73 ± 0.37	1.00 ± 0.09	0.87 ± 0.12
$\mu + jets$	1.24 ± 0.34	0.98 ± 0.34	0.97 ± 0.08	1.02 ± 0.09

Channel	F_{bb}, F_{cc}	F_c	F_{ll}	Overall normalization factor
$e+jets$	1.36 ± 0.27	0.88 ± 0.53	0.85 ± 0.16	0.65 ± 0.20
$\mu+jets$	1.43 ± 0.35	0.77 ± 0.38	0.79 ± 0.18	0.81 ± 0.17

Table 6.1 – Data-driven scale factors and uncertainties for the $W+jets$ samples on 2012 analysis [84].

6.2.3 Detector and reconstruction modeling

The uncertainties due to the non-perfect understanding of the detector and the reconstruction method are described in this section. The same methods have been used in both 2011 and 2012 analyses.

JET ENERGY SCALE

The Jet Energy Scale (JES) term represents one of the main contributions to the systematic uncertainty in jet-based analyses and its determination is a rather challenging task due to the difficult environment of hadron-hadron colliders. The jet energy scale depends on a variety of detectors and physic effects that includes the non-linearity in the calorimeter response, hardware problems of the detector and additional energy due to the underlying and pile-up event. The possible loss of energy during the jet reconstruction procedure is also included in this source of systematic uncertainty. In the 2012 the JES uncertainty for *large- R jets* is the main systematic contribution which estimation is still ongoing.

JET ENERGY RESOLUTION

The jet energy resolution (JER) uncertainty accounts for the effect due to the finite resolution on the jet energy estimation in the reconstruction phase; this value is obtained from MC simulations and validated using collision data. The impact of the uncertainty on the JER is evaluated by smearing jets energies

according to the systematic uncertainties on the resolution measurements performed on data [79].

B-TAGGING SCALE FACTOR

The corrections accounting for differences in the b-tagging efficiencies and mis-tag rates between data and simulation, are derived from data and parameterized as a function of p_T and η [80, 81]. The uncertainty on these corrections represents a source of systematic uncertainty evaluated by varying the scale factor by $\pm 1\sigma$ standard deviation.

LEPTON CONTRIBUTION

Electron and muon trigger, reconstruction and selection efficiencies are measured using data with the presence of W and Z bosons and have been incorporated as appropriate correction factors into the simulations. A similar procedure is also used to correct the lepton energy or momentum scale and resolution to match those observed in data.

MISSING TRANSVERSE ENERGY

The E_T^{miss} systematic uncertainty term has been evaluated on multi-jet simulation samples by varying both the amount of dead material used in detector simulation, by changing the parton shower models and by using different minimum bias tunes in order to simulate the underlying soft interactions.

LUMINOSTY

The 2011 data luminosity has been measured from Van Der Meer scans. Its uncertainty was estimated to be 3.7%. This uncertainty is applied on all non-normalized differential cross section measurements.

The uncertainty on the 2012 integrated luminosity resulted to be 2.8%, derived from a preliminary luminosity scale calibration [82] using beam-separation scans performed in November 2012.

6.3 Results of the 2011 data analysis

In order to obtain the $t\bar{t}$ event distributions, the SVD unfolding procedure has been applied to the data events passing the analysis selection after the subtraction of the background contributions described in Chapter 3. The migration matrices obtained from MC simulation, as well as the MC background distributions, are normalized to the data integrated luminosity by multiplying for the scale factor $\frac{\mathcal{L}_{data}}{\mathcal{L}_{MC}}$, where \mathcal{L}_{MC} is the integrated luminosity used to produce the MC sample considered. The resulting distributions are then multiplied by the bin-per-bin efficiencies ε_i in order to be comparable with the theoretical prediction evaluated at the *parton level*.

The evaluation of the differential cross section has been performed with respect to the top quark transverse momentum, the mass, the p_T and the rapidity of the $t\bar{t}$ system. In order to improve the precision of the measurement, the relative differential cross section, defined as normalized with respect to the integrated cross section, has been evaluated. In fact most of the systematic uncertainty contributions are cancelled in the cross section ratio $\frac{1}{\sigma} \frac{d\sigma}{dX}$ where X is the kinematic variable referred to the differential cross section. The remaining systematic contributions are the ones affecting only the distribution shape.

In each bin i , the cross section has been evaluated via the following equation

$$\sigma_i = \frac{e_i}{\mathcal{L} \cdot w_i} \quad 5.1$$

where e_i is the number of produced events after unfolding and efficiency correction and w_i is the bin width, both evaluated in the i^{th} bin; \mathcal{L} is the data sample luminosity. The integrated cross section has evaluated by adding all the events in a single bin including all the kinematic range and then repeating the analysis procedure already described. This procedure is preferred to the integration of the differential cross section in order to reduce the uncertainties.

The cross section results are compared with respect to the theoretical predictions from the MC@NLO, MCFM[83] and *ALPGEN* generators. The first two predictions are obtained from a full NLO matrix element calculation of the $pp \rightarrow t\bar{t}$ process, while the last one is generated with LO calculation of the $pp \rightarrow t\bar{t} + Njets$ process. The MCFM prediction is performed without the simulation of the parton showering. The agreement

between the results and the theoretical predictions can be seen from the ratio between the measured cross section distributions and the theoretical prediction considered; the ratio distributions are shown in the bottom part of the respective plots presented below. In the MCFM comparison the uncertainties are obtained by varying the factorization and renormalization scales $\mu = m_{top}$ in the range $[2\mu, \mu/2]$. In the comparison with MC@NLO and ALPGEN predictions, the uncertainties are calculated propagating the systematic uncertainties through unfolding, as described in Paragraph 5.4.5.

The various differential cross section results obtained during the 2011 analysis will be presented in the following paragraphs.

6.3.1 Differential cross section as a function of the top quark p_T : $\frac{1}{\sigma} \frac{d\sigma}{dp_T^t}$

The *parton level* differential cross section of the $t\bar{t}$ production as a function of the transverse momentum of the top quark $\frac{1}{\sigma} \frac{d\sigma}{dp_T^t}$ is shown in Figures 6.2 and 6.3 and in Table 6.2; results from electron channel, muon channel and the combination of the two are considered and compared with theoretical predictions. The cross section distributions, represented by the black points, are in good agreement with the theoretical predictions as can be seen from the ratio plots, even if in the highest p_T bins, the ALPGEN prediction is slightly higher. Moreover the ALPGEN prediction agrees slightly better with the data in both the electron and muon channels as for the overall normalization.

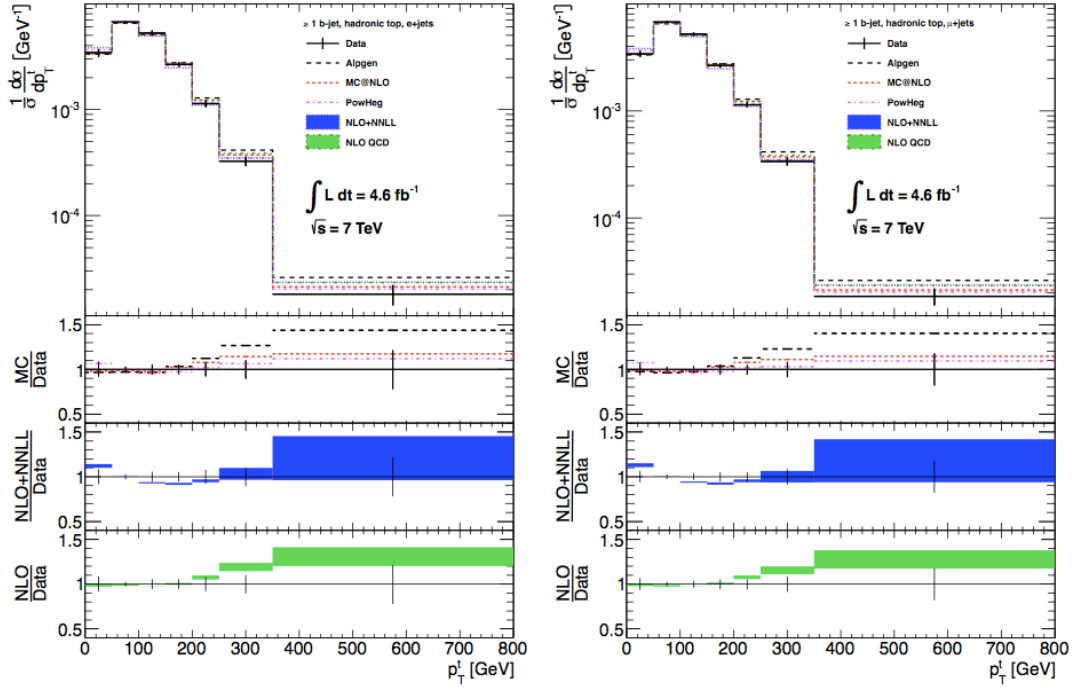


Figure 6.2 – Unfolded normalized differential cross section as function of the top quark p_T in $e^+ \text{jets}$ (left) and the $\mu^+ \text{jets}$ (right) channels. From top to bottom, the ratios of the Alpgen, MC@NLO and PowHeg predictions with the experimental results are displayed, as well as comparison to NLO+NNLL calculations and MCFM. The error bars on data points represent the combined statistical and systematic uncertainty on the measurement while the bands in the ratio plots denote theory scale variations. Alpgen was used to unfold the data.

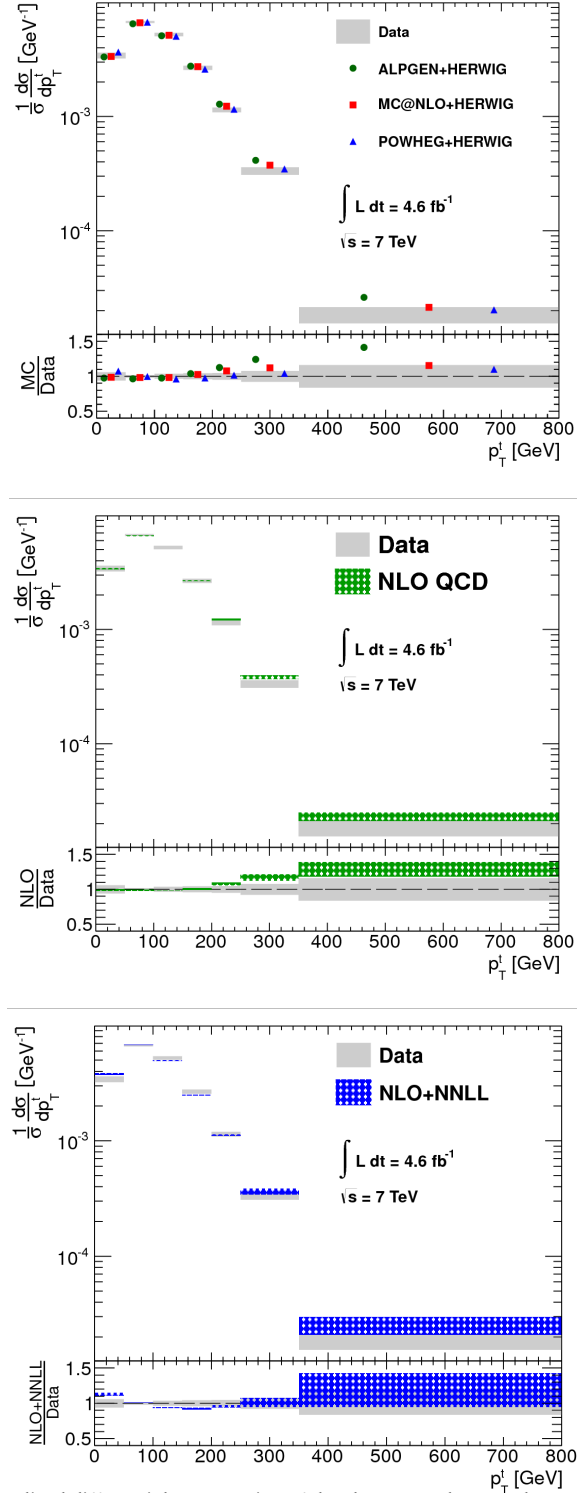


Figure 6.3 – Normalized differential cross section of the electron and muon channels combination as function of hadronic top quark p_T are compared with the Alpgen, MC@NLO and PowHeg MC predictions (top), with the NNLO prediction with MCFM (center) and with the NLO+NNLL result (bottom). The grey bands indicate the total uncertainty on the data while the coloured one in the middle and bottom plots represent the theoretical uncertainty on used μ scale and PDF.

6.3.2 Differential cross section as a function of

$$t\bar{t} \text{ mass: } \frac{1}{\sigma} \frac{d\sigma}{dm_{t\bar{t}}}$$

The *parton level* differential cross section of the $t\bar{t}$ production as a function of the mass of the $t\bar{t}$ system $\frac{1}{\sigma} \frac{d\sigma}{dm_{t\bar{t}}}$ is shown in Figure 6.4 and 6.5 and in Table 6.3; results from electron channel, muon channel and the combination of the two are considered. The cross section distributions, represented by the black points, are in good agreement with the theoretical prediction as can be seen from the ratio plots.

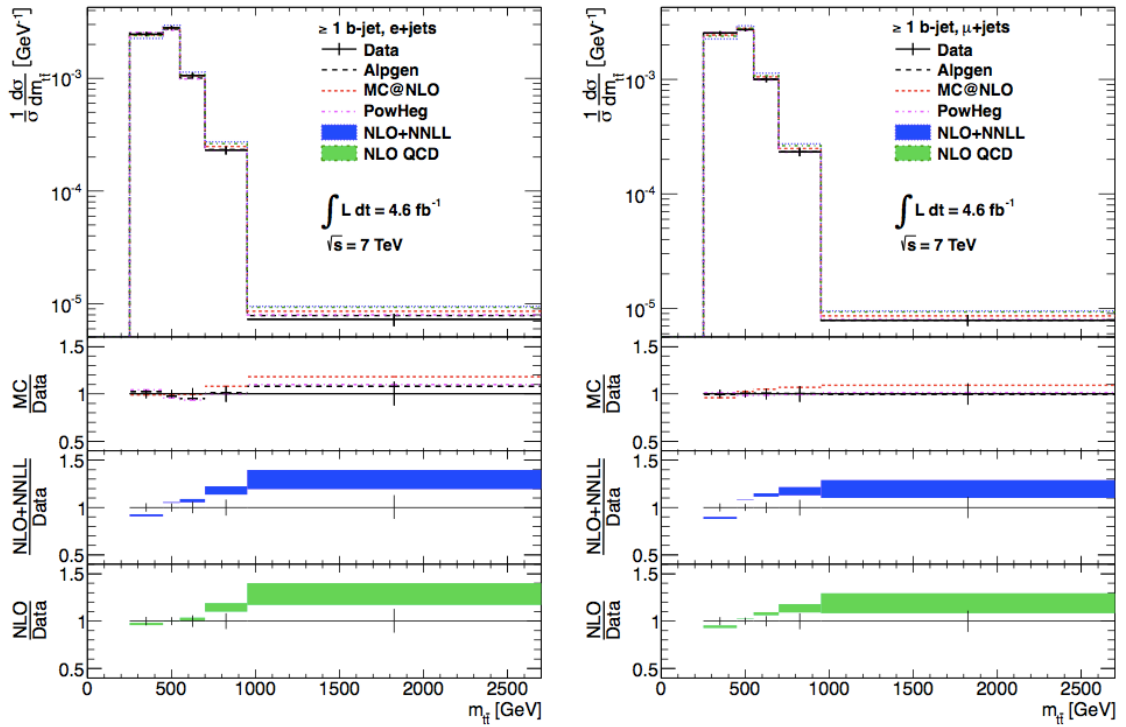


Figure 6.4 – Unfolded normalized differential cross section as function of $t\bar{t}$ mass in e +jets (left) and the μ +jets (right) channels. From top to bottom, the ratios of the Alpgen, MC@NLO and PowHeg predictions with the experimental results are displayed, as well as comparison to NLO+NNLL calculations and MCFM. The error bars on data points represent the combined statistical and systematic uncertainty on the measurement while the bands in the ratio plots denote theory scale variations.

Alpgen was used to unfold the data.

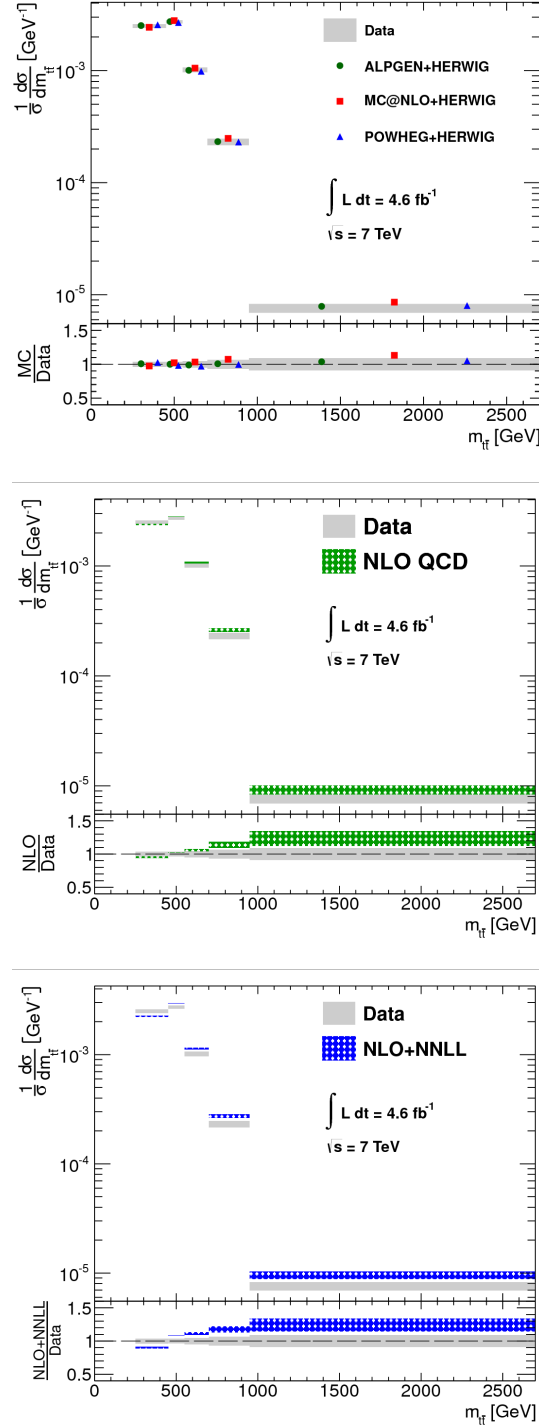


Figure 6.5 – Normalized differential cross section of the electron and muon channels combination as function of hadronic top quark p_T are compared with the Alpgen, MC@NLO and PowHeg MC predictions (top), with the NNLO prediction with MCFM (center) and with the NLO+NNLL result (bottom). The grey bands indicate the total uncertainty on the data while the coloured one in the central and bottom plots represent the theoretical uncertainty on used μ scale and PDF.

6.3.3 Differential cross section as a function of $t\bar{t}$ transverse momentum: $\frac{1}{\sigma} \frac{d\sigma}{dp_T^{t\bar{t}}}$

The *parton level* differential cross section of the $t\bar{t}$ production as a function of the transverse momentum of the $t\bar{t}$ system $\frac{1}{\sigma} \frac{d\sigma}{dp_T^{t\bar{t}}}$, is shown in Figure 6.6 and 6.7 and in Table 6.4; results from electron channel, muon channel and the combination of the two are considered. The cross section distributions, represented by the black points, well reproduce both the MC@NLO and *ALPGEN* theoretical predictions. On the contrary, they do not agree with the theoretical prediction from MCFM; in the first two bins the measured cross section is not correctly described within the uncertainties. The reason of that is the strong sensitivity of the $p_T^{t\bar{t}}$ spectra in the low p_T region to the additional radiation coming from non-perturbative parton shower that is not included in the calculation provided by MCFM.

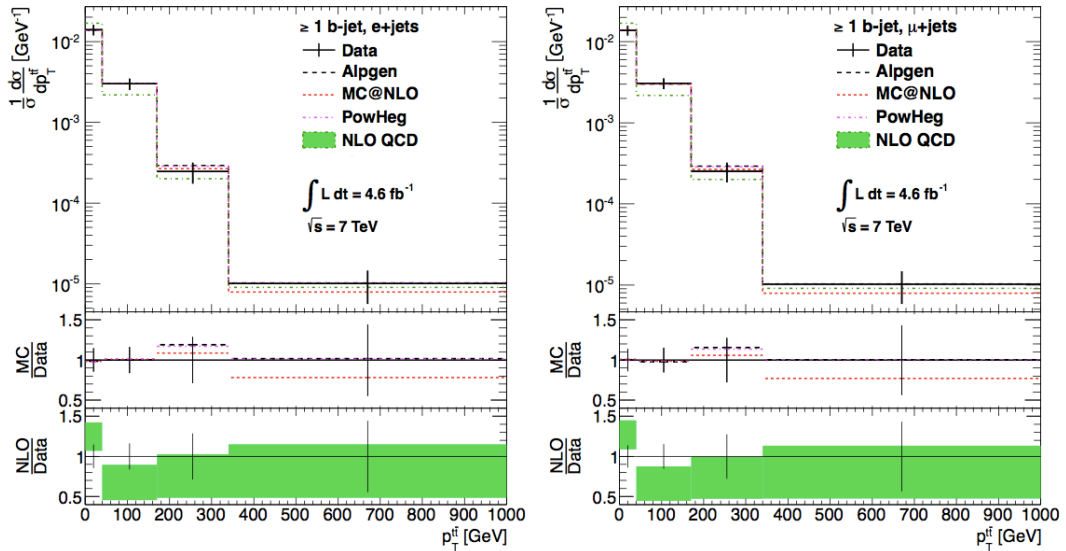


Figure 6.6 – Unfolded normalized differential cross section as function of $t\bar{t}$ transverse momentum in $e+jets$ (left) and the $\mu+jets$ (right) channels. From top to bottom, the ratios of the Alpgen, MC@NLO and PowHeg predictions with the experimental results are displayed, as well as comparison to MCFM. The error bars on data points represent the combined statistical and systematic uncertainty on the measurement while the bands in the ratio plots denote theory scale variations. Alpgen was used to unfold the data.

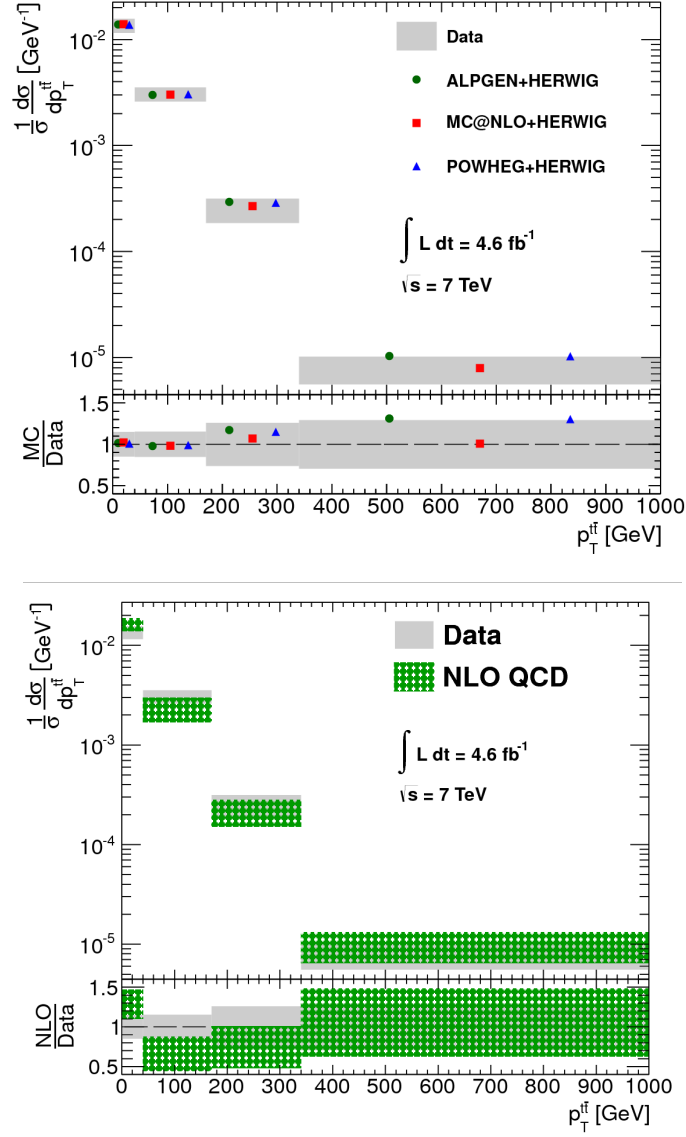


Figure 6.7 – Normalized differential cross section of the electron and muon channels combination as function of hadronic top quark p_T are compared with the Alpgen, MC@NLO and PowHeg MC predictions (top) and with the NLO prediction with MCFM (bottom). The grey bands indicate the total uncertainty on the data while the coloured one the bottom plots represent the theoretical uncertainty on used μ scale and PDF.

6.3.4 Differential cross section as a function of $t\bar{t}$ rapidity: $\frac{1}{\sigma} \frac{d\sigma}{dy_{t\bar{t}}}$

The *parton level* differential cross section of the $t\bar{t}$ production as a function of the rapidity of the $t\bar{t}$ system $\frac{1}{\sigma} \frac{d\sigma}{dy_{t\bar{t}}}$, is shown in Figure 6.8 and 6.9 and Table 6.5; results from electron channel, muon channel and the combination of the two are considered. The cross section distributions, represented by the black points, are in good agreement with the theoretical prediction as can be seen from the ratio plots.

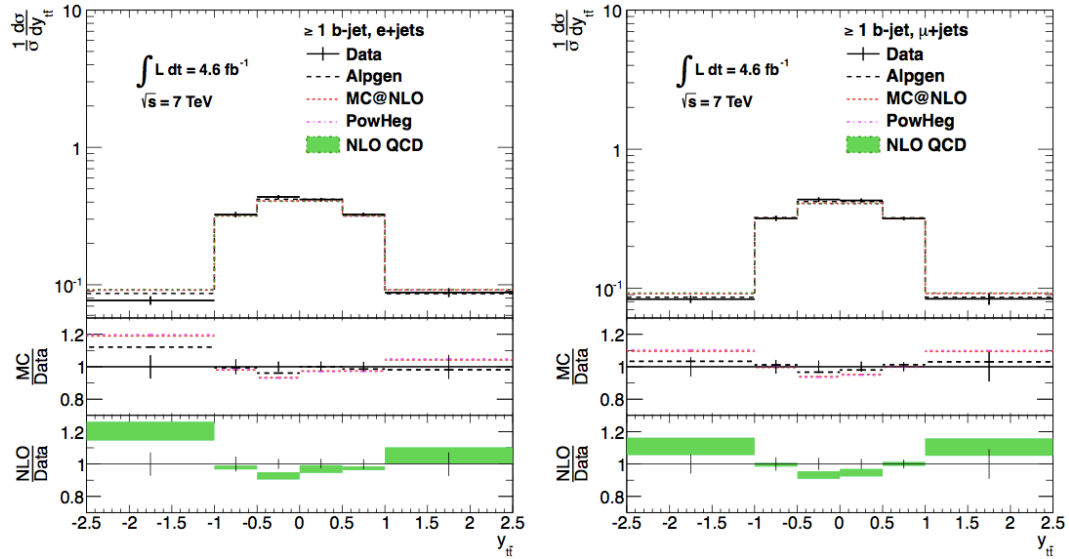


Figure 6.8 – Unfolded normalized differential cross section as function of $t\bar{t}$ rapidity in e -jets (left) and the μ -jets (right) channels. From top to bottom, the ratios of the Alpgen, MC@NLO and PowHeg predictions with the experimental results are displayed, as well as comparison to MCFM. The error bars on data points represent the combined statistical and systematic uncertainty on the measurement while the bands in the ratio plots denote theory scale variations. Alpgen was used to unfold the data.

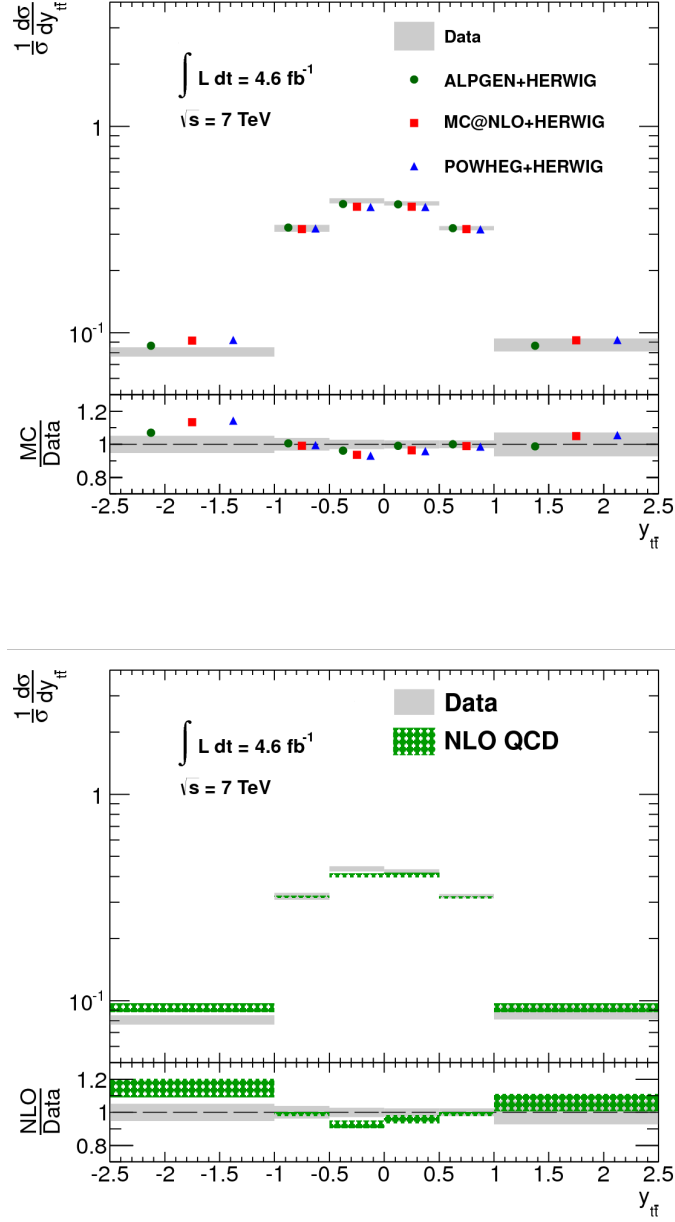


Figure 6.9 – Normalized differential cross section of the electron and muon channels combination as function of hadronic top quark p_T are compared with the Alpgen, MC@NLO and PowHeg MC predictions (top) and with the NNLO prediction with MCFM (bottom). The grey bands indicate the total uncertainty on the data while the coloured one in the bottom plot represent the theoretical uncertainty on used μ scale and PDF.

p_T^t [GeV]	$\frac{1}{\sigma} \frac{d\sigma}{dp_T^t} [10^{-3}]$	stat. [%]	syst. [%]
0 to 50	3.4 ± 0.1	± 2	± 4
50 to 100	6.7 ± 0.1	± 1	± 1
100 to 150	5.2 ± 0.1	± 2	± 2
150 to 200	2.66 ± 0.08	± 2	± 3
200 to 250	1.14 ± 0.04	± 2	± 3
250 to 350	0.33 ± 0.02	± 3	± 5
350 to 800	0.018 ± 0.002	± 6	± 10

Table 6.2 – Electron and muon combined normalized differential cross section as a function of the top quark p_T : values in each bin are presented with the total uncertainties. The statistical and systematic contribution is detailed in the last two columns.

$m_{t\bar{t}}$ [GeV]	$\frac{1}{\sigma} \frac{d\sigma}{dm_{t\bar{t}}} [10^{-3}]$	stat. [%]	syst. [%]
250 to 450	2.50 ± 0.08	± 1	± 3
450 to 550	2.73 ± 0.07	± 1	± 2
550 to 700	1.02 ± 0.04	± 2	± 4
700 to 950	0.23 ± 0.01	± 3	± 4
950 to 2700	0.0076 ± 0.0005	± 4	± 5

Table 6.3 – Electron and muon combined normalized differential cross section as a function of the $t\bar{t}$ mass: values in each bin are presented with the total uncertainties. The statistical and systematic contribution is detailed in the last two columns.

$p_T^{t\bar{t}}$ [GeV]	$\frac{1}{\sigma} \frac{d\sigma}{dp_T^{t\bar{t}}} [10^{-3}]$	stat. [%]	syst. [%]
0 to 40	14 ± 2	± 3	± 10
40 to 170	3.1 ± 0.4	± 2	± 10
170 to 340	0.25 ± 0.06	± 4	± 20
340 to 1000	0.008 ± 0.002	± 8	± 20

Table 6.4 – Electron and muon combined normalized differential cross section as a function of the $t\bar{t}$ transverse momentum: values in each bin are presented with the total uncertainties. The statistical and systematic contribution is detailed in the last two columns.

$y_{t\bar{t}}$	$\frac{1}{\sigma} \frac{d\sigma}{dy_{t\bar{t}}} [10^{-3}]$	stat. [%]	syst. [%]
-2.5 to -1.0	81 ± 3	± 2	± 3
-1.0 to -0.5	321 ± 9	± 1	± 3
-0.5 to 0.0	436 ± 9	± 1	± 2
0.0 to 0.5	423 ± 7	± 1	± 1
0.5 to 1.0	321 ± 5	± 1	± 1
1.0 to 2.5	87 ± 5	± 3	± 4

Table 6.5 – Electron and muon combined normalized differential cross section as a function of the $t\bar{t}$ rapidity: values in each bin are presented with the total uncertainties. The statistical and systematic contribution is detailed in the last two columns.

6.4 Results of the 2012 analysis

In order to obtain the $t\bar{t}$ event distributions, the SVD unfolding procedure is applied to data events passing the boosted selection analysis described in Paragraph 5.1.2, after the subtraction of the background contributions described in Chapter 3. The migration matrices obtained from MC simulation, as well as the MC background distributions, are normalized to the data integrated luminosity by multiplying for the scale factor $\frac{\mathcal{L}_{data}}{\mathcal{L}_{MC}}$, where \mathcal{L}_{MC} is the integrated luminosity used to produce the considered MC sample. The choice to measure the differential cross section at the *particle level* (see Paragraph 5.4.5) requires the definition of a fiducial phase space region via the application of a series of event cuts on the truth objects. The unfolding procedure should be only performed, by definition, to those events that have passed both the particle and the reconstructed selections. This implies to include, before unfolding, an additional acceptance scale factor in order to correct for reconstructed events that do not enter in fiducial region; this factor is evaluated by dividing the population of the MC events fulfilling both reconstructed and particle level selection by the population of MC events which pass at least the reconstructed one. The resulting unfolded distribution should then be multiplied by the bin-per-bin efficiencies ε_i in order to be comparable with the theoretical predictions also evaluated at the *particle level*. Efficiencies ε_i are defined as the ratio between the number of events included in the fiducial region and the event population fulfilling both particle and reconstruction selections.

The differential cross section as a function of the hadronic top quark transverse momentum $\frac{d\sigma}{dp_T^t}$ has been measured. In each bin, the cross section has been evaluated via the following equation

$$\sigma_i = \frac{e_i}{\mathcal{L} \cdot w_i} \quad 5.1$$

where e_i are respectively the number of events after unfolding and efficiency correction while w_i are the width of the i^{th} bin.

The cross section results are compared with respect to the theoretical predictions given by the *Powheg+Pythia* generator. The agreement between the results and the theoretical prediction can be seen from the ratio between the measured cross section distributions and the theoretical prediction ones; the ratio distributions are shown in the bottom part of the respective plots. The uncertainties are propagated through unfolding, as described in Paragraph 5.4.5.

The high- p_T $t\bar{t}$ pair production differential cross section as a function of the transverse momentum of the top quark $\frac{d\sigma}{dp_T^t}$ evaluated at the *particle level* is shown in Table 6.6 and Figure 6.10; electron and muon channel results are presented separately. The set of systematic uncertainties considered is not yet complete and in particular two sources which contribution strength have been evaluated as one the most significant from similar-selection analysis, still missing. For this reason, in the results presented in the following, only statistical uncertainties are considered. The contribution from the current set of systematic sources is presented in Appendix A. The cross section distributions, represented by the blue dots, are found to have a discrepancy of about 30%

with respect to the theoretical prediction coming from the Powheg+Pythia generator, as anticipated by the similar discrepancy in the data/MC comparison shown in Chapter 5. The cause of this disagreement is still under investigation; some attempts are ongoing to find the discrepancies source looking at the distribution behavior by changing the event selection parameters.

e+jets channel		$\mu + jets$ channel	
$p_T^t [GeV]$	$\frac{d\sigma_{t\bar{t}}}{dp_T^t} [fb^{-1}]$	$p_T^t [GeV]$	$\frac{d\sigma_{t\bar{t}}}{dp_T^t} [fb^{-1}]$
300 to 350	3.4 ± 0.2	300 to 350	5.6 ± 0.3
350 to 400	1.9 ± 0.1	350 to 400	3.45 ± 0.19
400 to 450	1.3 ± 0.1	400 to 450	2.02 ± 0.13
450 to 500	0.75 ± 0.08	450 to 500	1.28 ± 0.10
500 to 550	0.37 ± 0.05	500 to 550	0.68 ± 0.07
550 to 650	0.18 ± 0.02	550 to 650	0.27 ± 0.03
650 to 750	0.060 ± 0.013	650 to 750	0.10 ± 0.02
750 to 1200	0.006 ± 0.002	750 to 1200	0.010 ± 0.002

Table 6.6 – Unfolded differential cross section as function of the hadronic top quark transverse momentum in e+jets (top) and the μ +jets (bottom) channels. Values in each bin are presented with the sistematic uncertainties only.

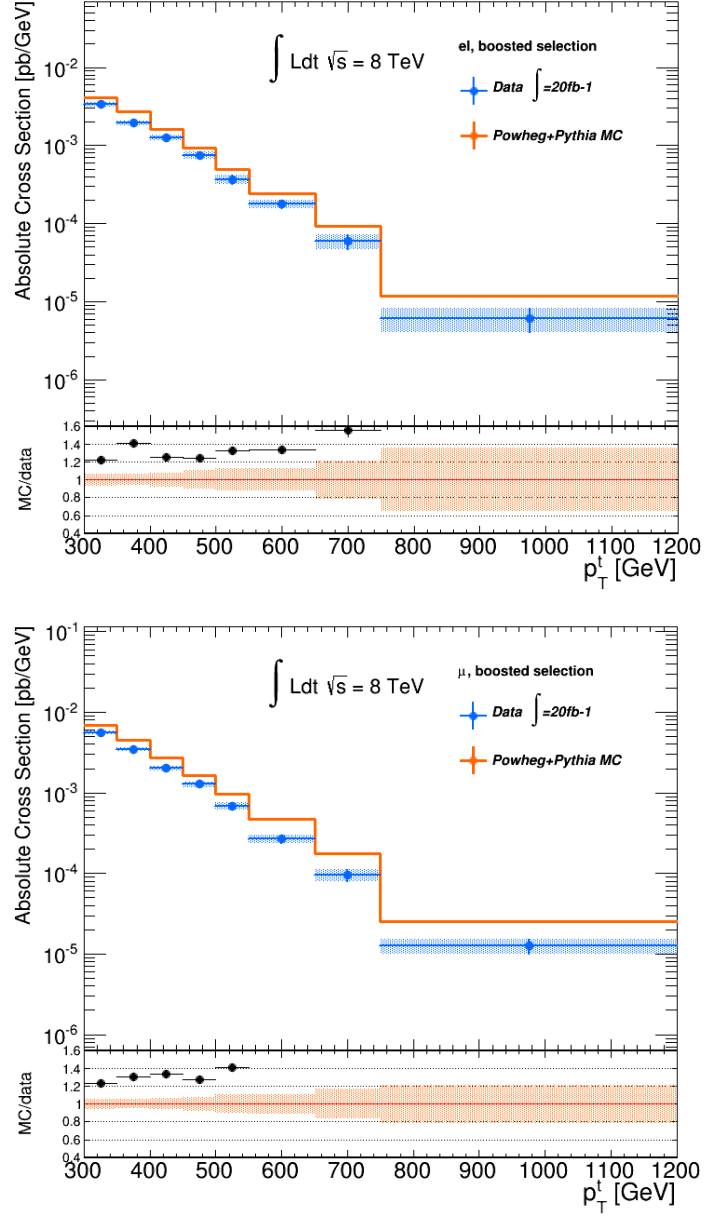


Figure 6.10 – Unfolded differential cross section as function of the hadronic top quark transverse momentum in $e+jets$ (top) and the $\mu+jets$ (bottom) channels. The results are compared with Powheg+Pythia predictions both in the distribution and in the ratio plots. The error bars represent the combined statistical and systematic uncertainty on the measurement.

Conclusion

The results of two analyses, both measuring the $t\bar{t}$ production differential cross section in the *lepton+jets* decay channel, have been described. The events considered have been collected at the LHC accelerator by the ATLAS experiment during pp collisions.

The first analysis has been performed with the full 2011 dataset corresponding to an integrated luminosity $\mathcal{L} = 4.6fb^{-1}$ at a center of mass energy $\sqrt{s} = 7TeV$. The backgrounds have been estimated via Monte Carlo simulation and by using *data*-driven techniques. The cut-based event selection requires one high p_T isolated lepton, the presence of significant missing transverse energy indicating the presence of a neutrino and at least four jets, one of which tagged as a b quark. A likelihood kinematic fit has been used in order to reconstruct the $t\bar{t}$ system. In order to further enhance the efficiency and purity of the reconstructed signal, a likelihood cut has been applied. The

single value decomposition (SVD) unfolding method has been applied in order to remove the acceptance and the resolution effects of the detector, giving the possibility to compare the results with the theoretical predictions, by unfolding data at the *parton level*. A detailed analysis of the sources of systematic uncertainties has been performed finding the Jet Energy Scale as the dominant one. The relative differential cross sections $\frac{1}{\sigma} \frac{d\sigma}{dX}$ have been measured as a function of the mass, p_T and rapidity of the $t\bar{t}$ system and of the p_T of the top quark. The results have been compared with NLO and NLO+NNLL theoretical calculations and MC generator predictions from Alpgen, MC@NLO and Powheg, finding a good agreement. In this measurement, I followed the whole analysis chain, specifically concentrating my efforts on the implementation of the unfolding techniques and the estimation of the systematic uncertainties. The results have been already validated by the ATLAS collaboration and a public conference note has been published[76]; a paper is on the way for publication too[95].

The second analysis presented exploited the full 2012 statistic ($\mathcal{L} = 20fb^{-1}$) collected at a center of mass energy $\sqrt{s} = 8TeV$ that was used in order to study the $t\bar{t}$ production cross section behavior at high top quark p_T . The event selection was very similar to the one used for the 2011 data analysis; the most relevant change was the request of at least one *large- R jet* with specifically tuned cuts on jet substructure variables. This replaced the usual request of at least four *small- R jets* applied in the 2011 data analysis. Only preliminary studies on the more sophisticated template overlap tagging algorithm (TOM) have been shown because its application to the analysis is still under investigation. The backgrounds have been evaluated both from MC simulation

and *data-driven* techniques, as in the W+jets channel. The QCD background contribution has been estimated to be about 1%. The selected events have been unfolded using the SVD method to the *particle level* that allows an easier and more model-independent result comparison with respect to the unfolding at *parton level*. The effect of the main systematic uncertainties has been estimated. The final differential cross section result as a function of the hadronic top $p_T \left(\frac{d\sigma}{dp_T^t} \right)$ has been compared with the MC prediction from the *Powheg+Pythia* generator, founding a discrepancy of the order of about 20–30%. The source of this discrepancy is still under investigation and some attempts are ongoing to understand the disagreement by changing the event selection requirements. I am currently the main code developer and analyzer for this measurements and I am performing all the analysis steps from the selection to the background estimation and the unfolding process. This analysis results are collected in an ATLAS note on preparation of which I am one of the editors; this will evolve in a public paper for this summer. In the future, I will continue my analysis work on these items and I will also measure the cross section dependences from more top quark and $t\bar{t}$ kinematic variables beside the estimation of the total set of systematics involved, further investigations about the origin of the data/MC discrepancy and the estimation of the differential cross section at the *parton level*. I will furthermore include in the event selection chain the TOM tagger in order to enhance the event selection.

Appendix A

Systematic uncertainty tables

For the sake of compactness, the many systematic sources affecting the 2011 data analysis differential cross section evaluated as a function of $t\bar{t}$ variables have been divided in five groups depending on the object affected:

- jets;
- leptons;
- energy scale (affecting mainly the E/ T measurement);
- fake-lepton and W background, containing the uncertainties on the data driven background estimation;

- Monte Carlo generation, theory, IFSR and PDF, affecting both signal and background modeling.

The uncertainty on the luminosity isn't included because, since this measurements are about the normalized cross section $\frac{1}{\sigma} \frac{d\sigma}{dX}$, the luminosity induced on the fluctuations on the $\frac{1}{\sigma}$ systematically cancel with the ones on the $d\sigma$ term.

In tables A.7 and A.8 the differential cross section results are presented with the contribution given by the limited sample of systematic considered and detailed in Tables A.5 and A.6

$\frac{d\sigma}{dp_T}$ Uncertainties [%]	0 to 50	50 to 100	100 to 150	150 to 200	200 to 250	250 to 350	350 to 800
JES Uncertainty in Background	0.21	0.27	0.22	0.13	0.08	0.17	0.24
<i>b</i> -tagging Efficiency	0.95	1.15	1.83	2.18	2.93	3.88	4.75
Fragmentation	0.49	0.12	0.33	0.07	0.55	0.57	3.00
IFSR	0.10	0.14	0.14	0.96	0.46	0.04	0.03
<i>b</i> -quark Jets (JES)	0.10	0.00	0.21	0.16	0.08	0.05	0.04
Closeby Jets (JES)	0.02	0.02	0.19	0.34	0.33	0.54	0.77
Effective Detector NP Set 1 (JES)	0.06	0.00	0.05	0.16	0.16	0.18	0.19
Effective Model NP Set 1 (JES)	0.37	0.29	0.27	0.33	0.24	0.13	0.03
Effective Stat. NP Set 1 (JES)	0.27	0.15	0.07	0.06	0.05	0.01	0.00
η -Intercalibration (JES)	0.01	0.04	0.11	0.17	0.11	0.07	0.05
Flavor Response (JES)	0.24	0.12	0.04	0.05	0.05	0.03	0.03
Kin. Fit Likelihood SF	0.08	0.08	0.07	0.06	0.07	0.09	0.10
<i>l</i> ID SF	0.07	0.00	0.16	0.07	0.11	0.16	0.08
<i>l</i> Reco. SF	0.01	0.00	0.03	0.01	0.02	0.03	0.01
<i>l</i> Trigger SF	0.08	0.17	0.40	0.05	0.03	0.03	0.08
Luminosity	0.27	0.24	0.24	0.21	0.23	0.28	0.32
MC Generator	6.98	9.67	8.61	10.15	8.21	5.14	0.17
PDF	0.37	0.33	0.30	0.21	0.16	0.11	0.13
Multijet Normalization	0.19	0.00	0.01	0.00	0.01	0.11	0.13
Multijet Shape	0.01	0.00	0.01	0.01	0.01	0.00	0.01
W+jets bbcc	0.57	0.22	0.08	0.07	0.06	0.03	0.02
W+jets c4	0.05	0.04	0.03	0.00	0.00	0.00	0.00
W+Jets c5	0.05	0.04	0.05	0.00	0.00	0.00	0.00
W+jets Charge Asymm.	0.48	0.18	0.06	0.05	0.06	0.04	0.07

Table A.1 – Percentage contribution of each uncertainty to the total combined uncertainty calculated using the BLUE method in the 2011 data analysis. The uncertainties are symmetrized for hadronic $top p_T$. These symmetrized uncertainties are calculated by averaging the asymmetric up and down components of each uncertainty. This is an approximation, however, as the asymmetry in most uncertainties is not extreme, it is not an unreasonable approximation. This is for the absolute spectrum.

$1/\sigma d\sigma_{t\bar{t}}/dm_{t\bar{t}}$	$m_{t\bar{t}}, l + jets$				
Uncertainty [%]	250 – 450	450 – 550	550 – 700	700 – 960	960 – 2700
Total	5.3/-5.3	4.0/-4.0	5.3/-5.4	7.4/-7.6	12/-12
Stat. only	± 1.8	± 1.6	± 2.1	± 3.5	± 5.7
Syst. only	5.0/-4.9	3.6/-3.7	4.9/-5.0	6.5/-6.8	10/-11
Jets	2.7/-2.6	0.9/-1.0	3.3/-3.5	5.1/-5.4	6.8/-7.2
Leptons	0.05/-0.04	0/-0.11	0.05/-0.07	0.2/-0.12	0.5/-0.3
Energy Scale	0.06/-0.07	0.15/-0.14	-0.014	0.19/-0.5	0.3/-1.1
Fake-lepton and W back-ground	4.1/-4.1	3.5/-3.5	3.3/-3.3	3.6/-3.6	4.1/-4.2
Monte Carlo gen., theory, IFSR and PDF	0.9/-0.9	0.4/-0.4	1.3/-1.3	1.8/-1.8	6.6/-6.6

Table A.2 – Contribution of each systematic uncertainty grouped in some main categories to the total $(1/\sigma) \cdot d\sigma/dm_{t\bar{t}}$ spectrum as percentage of the normalized cross-section, for the combination of the $e+jets$ and $\mu+jets$ channels in the 2011 data analysis.

$1/\sigma d\sigma_{t\bar{t}}/dp_{T,t\bar{t}}$	$p_{T,t\bar{t}}, l + jets$			
Uncertainty [%]	0 – 40	40 – 170	170 – 340	340 – 1200
Total	13/-12	8.3/-9.0	19/-20	25/-25
Stat. only	± 2.3	± 2.0	± 4.8	± 8.7
Syst. only	12/-12	8.0/-8.8	19/-19	23/-24
Jets	6.7/-5.6	3.5/-4.8	9.7/-11	13/-14
Leptons	0.3/-0.19	0.09/-0.3	0.16/-0.06	0.6/-0
Energy Scale	2.8/-2.2	1.7/-2.2	1.7/-1.9	1.7/-1.1
Fake-lepton and W background	4.0/-4.0	3.5/-3.5	4.2/-4.2	4.9/-5.1
Monte Carlo gen., theory, IFSR and PDF	9.2/-9.2	6.1/-6.1	15/-15	18/-18

Table A.3 – Contribution of each systematic uncertainty grouped in some main categories to the total $(1/\sigma) \cdot d\sigma/dp_T^{t\bar{t}}$ spectrum as percentage of the normalized cross-section, for the combination of the $e+jets$ and $\mu+jets$ channels in the 2011 data analysis.

$1/\sigma d\sigma_{t\bar{t}}/dy_{t\bar{t}}$	$y_{t\bar{t}}, l + jets$					
Uncertainty [%]	-2.5 – -1	-1 – -0.5	-0.5 – 0	0 – 0.5	0.5 – 1	1 – 2.5
Total	5.4/-5.9	4.3/-4.3	4.2/-4.1	4.2/-4.2	4.2/-4.2	5.4/-5.3
Stat. only	± 3.1	± 1.9	± 1.8	± 1.9	± 2.0	± 3.2
Syst. only	4.5/-5.0	3.8/-3.8	3.8/-3.7	3.7/-3.8	3.7/-3.7	4.3/-4.2
Jets	1.8/-2.8	0.3/-0.4	1.0/-0.7	0.7/-0.8	0.3/-0.3	1.5/-1.2
Leptons	0/-0.2	0.06/-0.018	0.03/-0.13	0.06/-0.09	0.012/-0.06	0.18/-0.06
Energy Scale	0.2/-0.3	0.05/-0.02	0.06/-0.05	0.09/-0.05	+0.2	+0.08
Fake-lepton and W background	4.1/-4.1	3.7/-3.7	3.6/-3.6	3.6/-3.6	3.7/-3.7	3.8/-3.8
Monte Carlo gen., theory, IFSR and PDF	0.5/-0.5	0.8/-0.8	0.3/-0.3	0.5/-0.5	0.2/-0.2	1.6/-1.6

Table A.4 – Contribution of each systematic uncertainty grouped in some main categories to the total $(1/\sigma) \cdot d\sigma/dy_{t\bar{t}}$ spectrum as percentage of the normalized cross-section, for the combination of the $e+jets$ and $\mu+jets$ channels in the 2011 data analysis.

	$\frac{d\sigma_{t\bar{t}}}{dp_T^t}$							
Uncertainty [%]	300 to 350	350 to 400	400 to 450	450 to 500	500 to 550	550 to 650	650 to 750	750 to 1200
Total	6.92/-6.83	6.21/-6.11	7.89/-7.84	10.23/-10.29	12.53/-12.58	12.11/-12.14	21.53/-21.32	35.30/-35.18
Stat. only[%]	± 6.66	± 5.90	± 7.76	± 10.19	± 12.46	± 11.91	± 20.97	± 33.69
Syst. only[%]	1.89/-1.54	1.91/-1.58	1.44/-1.09	0.87/-1.42	1.27/-1.68	2.16/-2.34	4.91/-3.85	10.51/-10.11
BJesUnc	0.12/-0.42	0.46/-0.46	0.38/-0.38	-0.12/-0.12	0.17/-0.59	0.55/-0.53	0.71/-0.10	1.70/-1.70
EtaIntercalibrationModel	0.25/-0.50	0.36/-0.14	0.79/-0.13	0.25/-0.19	0.12/-0.12	0.68/-0.68	1.37/-1.37	3.42/-3.42
EtaIntercalibrationStatMetho	0.06/-0.12	0.16/-0.16	0.19/-0.19	0.25/-0.21	0.12/-0.10	0.13/-0.23	0.22/-0.07	0.75/-0.75
JesEffectiveDet1	0.13/-0.30	0.31/-0.08	0.18/-0.18	0.01/-0.11	0.19/-0.04	0.25/-0.02	0.50/-0.50	1.64/-1.64
JesEffectiveDet2	0.05/-0.07	0.14/-0.14	0.10/-0.10	0.07/-0.22	0.10/-0.16	0.23/-0.25	0.62/-0.27	1.62/-0.16
JesEffectiveDet3	0.02/-0.00	0.19/-0.19	0.02/-0.02	-0.04/-0.04	0.15/-0.15	0.36/-0.36	0.18/-0.18	0.21/-0.01
JesEffectiveMix1	0.05/-0.08	0.19/-0.19	0.10/-0.00	0.01/-0.15	0.23/-0.23	0.25/-0.25	0.84/-0.10	2.15/-0.22
JesEffectiveMix2	0.05/-0.03	0.08/-0.08	0.02/-0.02	0.01/-0.03	0.03/-0.02	0.11/-0.11	0.10/-0.10	0.23/-0.02
JesEffectiveModel1	0.80/-0.69	0.56/-0.50	0.29/-0.27	0.13/-0.28	0.04/-0.38	0.26/-0.26	1.10/-1.10	2.63/-2.63
JesEffectiveModel2	-0.03/-0.03	0.26/-0.07	0.15/-0.15	-0.01/-0.01	0.13/-0.05	0.50/-0.50	0.51/-0.51	1.38/-1.38
JesEffectiveModel3	0.04/-0.03	0.24/-0.24	0.14/-0.14	0.02/-0.22	0.09/-0.25	0.07/-0.06	0.28/-0.28	0.85/-0.85
JesEffectiveModel4	0.09/-0.08	0.17/-0.17	0.03/-0.06	-0.07/-0.07	0.16/-0.11	0.23/-0.23	0.16/-0.16	0.20/-0.20
JesEffectiveStat1	0.06/-0.06	0.15/-0.15	0.06/-0.03	0.21/-0.11	0.14/-0.14	0.24/-0.23	0.22/-0.22	0.63/-0.63
JesEffectiveStat2	0.03/-0.03	0.12/-0.01	0.06/-0.06	0.02/-0.10	-0.01/-0.01	0.17/-0.17	0.73/-0.73	2.11/-2.11
JesEffectiveStat3	0.07/-0.16	0.13/-0.13	0.06/-0.06	0.03/-0.20	0.16/-0.22	0.20/-0.20	0.37/-0.37	0.93/-0.05
PileupOffsetM	-0.03/-0.03	0.09/-0.09	0.07/-0.04	0.20/-0.29	0.41/-0.16	0.24/-0.24	0.41/-0.41	0.47/-0.04
PileupOffsetNPV	0.18/-0.13	0.18/-0.24	0.03/-0.24	0.20/-0.40	0.11/-0.34	0.21/-0.01	0.87/-0.87	2.08/-2.08
PileupPt	-0.00/-0.00	0.12/-0.12	0.13/-0.13	-0.00/-0.00	0.09/-0.00	0.25/-0.25	0.40/-0.40	0.86/-0.86
PileupRho	0.31/-0.50	0.28/-0.20	0.16/-0.05	0.18/-0.33	-0.12/-0.12	0.02/-0.11	0.62/-0.62	1.71/-1.71
SinglePart	0.03/-0.02	0.13/-0.13	-0.00/-0.00	0.01/-0.01	0.07/-0.04	-0.07/-0.07	-0.07/-0.07	-0.01/-0.01
eer	0.11/-0.15	0.22/-0.22	0.40/-0.03	0.21/-0.30	0.47/-0.47	0.63/-0.63	0.77/-0.77	1.51/-1.51
ees	0.38/-0.39	0.36/-0.36	0.54/-0.38	0.33/-0.48	0.22/-0.60	0.71/-1.28	1.64/-0.63	2.58/-2.58
flavorcomp	0.67/-0.12	0.57/-0.38	0.00/-0.13	-0.05/-0.05	0.24/-0.15	0.73/-0.73	2.06/-2.06	4.26/-4.26
flavorresponse	0.44/-0.25	0.44/-0.27	0.04/-0.12	-0.05/-0.05	0.13/-0.19	0.43/-0.43	1.21/-1.21	2.84/-2.84
jes	1.30/-0.79	1.27/-1.05	0.49/-0.47	0.30/-0.46	0.47/-0.84	0.80/-0.05	1.26/-1.26	3.65/-3.65
lumi	0.08/-0.08	0.08/-0.08	0.08/-0.08	0.08/-0.08	0.08/-0.08	0.08/-0.08	0.08/-0.08	0.08/-0.08
ressoft	-0.10/-0.10	-0.24/-0.24	0.09/-0.18	0.42/-0.75	-0.57/-0.57	-0.44/-0.44	0.78/-0.07	1.39/-1.39
scsoft	0.40/-0.38	0.25/-0.25	0.56/-0.56	0.11/-0.43	-0.50/-0.50	0.67/-0.90	2.31/-0.71	3.38/-3.38
wjetNorm	0.02/-0.01	0.02/-0.02	0.01/-0.01	0.02/-0.02	0.02/-0.02	0.01/-0.01	0.02/-0.02	0.04/-0.04
wjetsFlavWc	0.01/-0.01	0.01/-0.01	0.01/-0.01	0.01/-0.01	0.01/-0.01	0.00/-0.00	0.01/-0.01	0.01/-0.01
wjetsFlavWhh	0.01/-0.01	0.02/-0.02	0.02/-0.02	0.03/-0.03	0.03/-0.03	0.02/-0.02	0.03/-0.03	0.07/-0.07

Table A.5 – Contribution of each systematic uncertainty considered in the 2012 data analysis to the $d\sigma/dp_T^t$ spectrum as percentage of the normalized cross-section, for the e^+jets channel.

	$\frac{d\sigma_{\ell\ell}}{dp_T^{\ell}}$							
Uncertainty [%]	300 to 350	350 to 400	400 to 450	450 to 500	500 to 550	550 to 650	650 to 750	750 to 1200
Total	5.63/-5.65	5.51/-5.62	6.54/-6.60	7.90/-7.90	10.18/-10.17	10.89/-10.90	16.60/-16.61	21.37/-21.37
Stat. only[%]	± 5.49	± 5.38	± 6.50	± 7.83	± 10.13	± 10.81	± 16.54	± 21.29
Syst. only[%]	1.24/-1.31	1.19/-1.61	0.74/-1.17	1.04/-1.06	1.02/-0.84	1.31/-1.44	1.42/-1.48	1.79/-1.77
BJesUnc	0.22/-0.25	0.07/-0.45	-0.15/-0.15	-0.06/-0.06	0.13/-0.01	0.08/-0.08	0.06/-0.06	0.05/-0.13
EtaIntercalibrationStatMetho	0.16/-0.16	-0.14/-0.14	-0.07/-0.07	0.13/-0.32	0.18/-0.40	0.12/-0.24	0.15/-0.00	0.34/-0.22
JesEffectiveDet1	0.20/-0.18	-0.31/-0.31	-0.05/-0.05	-0.06/-0.06	0.03/-0.28	-0.14/-0.14	0.04/-0.23	0.19/-0.35
JesEffectiveDet2	0.06/-0.01	-0.23/-0.23	-0.18/-0.18	-0.24/-0.24	-0.01/-0.01	0.34/-0.13	0.20/-0.15	-0.08/-0.08
JesEffectiveDet3	0.06/-0.07	-0.24/-0.24	-0.22/-0.22	-0.12/-0.12	0.16/-0.16	0.21/-0.21	0.16/-0.16	0.18/-0.14
JesEffectiveMix1	0.15/-0.03	-0.35/-0.35	-0.00/-0.00	0.25/-0.25	0.17/-0.17	0.29/-0.29	0.27/-0.27	0.32/-0.27
JesEffectiveMix2	0.22/-0.22	-0.21/-0.21	-0.25/-0.25	-0.06/-0.06	-0.04/-0.04	0.16/-0.02	0.23/-0.23	0.16/-0.16
JesEffectiveModel1	0.19/-0.70	-0.27/-0.27	-0.14/-0.14	0.25/-0.16	0.47/-0.17	-0.10/-0.10	-0.53/-0.53	-0.90/-0.90
JesEffectiveModel2	0.11/-0.11	-0.05/-0.05	0.02/-0.02	0.28/-0.28	-0.04/-0.04	-0.04/-0.04	0.03/-0.06	-0.08/-0.08
JesEffectiveModel3	0.02/-0.01	-0.25/-0.25	-0.09/-0.09	-0.03/-0.03	0.12/-0.12	0.25/-0.05	0.22/-0.19	0.18/-0.34
JesEffectiveModel4	0.14/-0.14	-0.29/-0.29	-0.13/-0.13	0.16/-0.07	0.26/-0.26	0.17/-0.17	-0.02/-0.02	-0.07/-0.07
JesEffectiveStat1	0.27/-0.27	-0.18/-0.18	-0.10/-0.10	-0.04/-0.04	0.38/-0.00	0.72/-0.72	0.47/-0.47	0.34/-0.34
JesEffectiveStat2	-0.01/-0.01	0.25/-0.25	0.15/-0.15	0.09/-0.13	-0.06/-0.06	-0.10/-0.10	-0.03/-0.03	0.22/-0.29
JesEffectiveStat3	0.21/-0.36	-0.15/-0.15	-0.00/-0.00	0.06/-0.27	-0.02/-0.02	0.19/-0.32	0.13/-0.04	0.13/-0.11
PileupOffsetM	0.03/-0.29	-0.18/-0.18	-0.08/-0.08	-0.14/-0.14	0.00/-0.08	0.25/-0.25	0.54/-0.01	0.77/-0.11
PileupOffsetNPV	0.02/-0.11	-0.27/-0.27	0.00/-0.24	0.04/-0.01	0.12/-0.12	0.16/-0.02	0.06/-0.04	-0.02/-0.02
PileupPt	0.10/-0.10	-0.28/-0.28	-0.09/-0.09	-0.01/-0.01	-0.10/-0.10	-0.08/-0.08	0.08/-0.02	0.16/-0.16
PileupRho	0.18/-0.22	0.02/-0.31	-0.09/-0.09	-0.08/-0.08	-0.05/-0.05	0.16/-0.39	0.24/-0.24	0.50/-0.50
SinglePart	0.07/-0.08	-0.07/-0.07	0.08/-0.16	0.02/-0.32	0.18/-0.11	0.40/-0.40	0.33/-0.33	0.21/-0.21
flavorcomp	0.31/-0.35	-0.22/-0.22	-0.13/-0.13	-0.13/-0.13	0.00/-0.18	-0.06/-0.06	0.38/-0.48	0.56/-0.92
flavorresponse	0.25/-0.39	-0.11/-0.11	-0.14/-0.14	0.02/-0.15	0.04/-0.04	-0.01/-0.01	-0.05/-0.05	0.01/-0.24
jcs	0.93/-0.57	0.22/-0.96	0.38/-0.74	0.80/-0.34	0.55/-0.32	0.05/-0.21	0.02/-0.50	-0.12/-0.12
lumi	0.08/-0.08	0.08/-0.08	0.08/-0.08	0.08/-0.08	0.08/-0.08	0.08/-0.08	0.08/-0.08	0.08/-0.08
mui	0.08/-0.02	-0.08/-0.08	-0.11/-0.11	-0.07/-0.07	-0.10/-0.10	0.06/-0.15	0.22/-0.22	0.23/-0.23
mums	0.18/-0.01	-0.13/-0.13	-0.08/-0.08	0.09/-0.26	0.14/-0.16	0.45/-0.45	0.55/-0.55	0.46/-0.46
musc	0.17/-0.05	-0.19/-0.19	0.06/-0.07	0.10/-0.10	0.13/-0.13	0.21/-0.21	0.37/-0.37	0.38/-0.16
ressoft	0.03/-0.18	-0.01/-0.01	0.10/-0.59	0.16/-0.55	0.24/-0.10	0.39/-0.46	0.03/-0.35	-0.11/-0.11
scsoft	0.04/-0.07	-0.57/-0.57	-0.21/-0.21	-0.02/-0.02	0.08/-0.21	0.17/-0.51	0.31/-0.35	0.38/-0.29
wjetNorm	0.02/-0.02	0.02/-0.02	0.03/-0.03	0.01/-0.01	0.01/-0.01	0.04/-0.04	0.04/-0.04	0.07/-0.07
wjetsFlavWc	0.00/-0.00	0.00/-0.00	0.00/-0.00	0.00/-0.00	0.00/-0.00	0.01/-0.01	0.02/-0.02	0.03/-0.03
wjetsFlavWhh	0.03/-0.03	0.04/-0.04	0.04/-0.04	0.03/-0.03	0.02/-0.02	0.06/-0.06	0.06/-0.06	0.11/-0.11

Table A.6 – Contribution of each systematic uncertainty considered in the 2012 data analysis to the $d\sigma/dp_T^{\ell}$ spectrum as percentage of the normalized cross-section, for μ -jets channel.

e+jets channel			
$p_T^t [GeV]$	$\frac{d\sigma_{t\bar{t}}}{dp_T^t} [fb^{-1}]$	stat. [%]	syst. [%]
300 to 350	3.36+0.23/-0.23	± 6	2/-2
350 to 400	1.9+0.1/-0.1	± 6	1.9/-1.5
400 to 450	1.3+0.1/-0.1	± 7	1.4/-1.1
450 to 500	0.75+0.08/-0.08	± 10	0.9/-1.4
500 to 550	0.37+0.05/-0.05	± 12	1.3/-1.7
550 to 650	0.18+0.02/-0.02	± 11	2/-2
650 to 750	0.060+0.013/-0.013	± 21	5/-4
750 to 1200	0.006+0.002/-0.002	± 34	11/-10

Table A.7 – Unfolded differential cross section as function of the hadronic top quark transverse momentum in μ +jets (top) channels. Values in each bin are presented with the total uncertainties. The statistical and systematic contribution is detailed in the last two columns.

$\mu + jets$ channel			
$p_T^t [GeV]$	$\frac{d\sigma_{t\bar{t}}}{dp_T^t} [fb^{-1}]$	stat. [%]	syst. [%]
300 to 350	5.6+0.3/-0.3	± 5	1.2/-1.3
350 to 400	3.45+0.19/-0.19	± 5	1.2/-1.6
400 to 450	2.02+0.13/-0.13	± 6	0.7/-1.2
450 to 500	1.28+0.10/-0.10	± 8	1.0/-1.1
500 to 550	0.68+0.07/-0.07	± 10	1.0/-0.8
550 to 650	0.27+0.03/-0.03	± 11	1.3/-1.4
650 to 750	0.10+0.02/-0.02	± 16	1.4/-1.5
750 to 1200	0.010+0.005/-0.002	± 21	1.8/-1.8

Table A.8 – Unfolded differential cross section as function of the hadronic top quark transverse momentum in e+jets (top) channels. Values in each bin are presented with the total uncertainties. The statistical and systematic contribution is detailed in the last two columns.

Appendix B

Systematic uncertainty tables

In a huge collaboration as ATLAS the transfer knowledge is a key point for an efficient teamwork in order to share tools for common tasks. This is possible only through a code with a friendly user interface and a high compatibility and flexibility.

On this prospective I have designed the analysis infrastructure used for the 2012 analysis called BoostedRealm: it is fully integrated in the most common ATLAS top analysis infrastructure named TopRootCore (TRC) that is derived and inherits from the most general RootCore ATLAS framework. The inclusion in TRC guarantees to easy maintain corrections and scale factors up to date and gives the possibility to use several tools exploiting for different analysis tasks already implemented by other groups.

The BoostedRealm framework has a modular structure trough a strong hierarchical coding. The different analysis steps, in order to simplify the managing and

to be as much user friendly as possible, are controlled via bash scripts and parameter files. A TRC package called TopD3PDBoosted, of which I am one of the developers, is included in the framework and deals with the events correction, selection and reconstruction; it is central validated from the ATLAS top working-group. The package is currently used in our $t\bar{t}$ differential cross section measurement as well as in the $t\bar{t}$ resonances search. The BoostedRealm framework also includes some other TRC independent programs dealing with the main analysis tasks as unfolding and systematic managing.

The principal programs needed during the analysis process are described below: D3PD2MiniSLBoosted (in the TopD3PDBoosted package) for event selection, HistoMakerMasterBase for histogram production and objects reconstruction the unfolding package for unfolding and systematic managing. The naming convention adopted is also described for its importance in the automation of the analysis procedure that allows the programs to identify histograms and file type by their name. In the following a quite technical description of the whole framework infrastructure is given.

D3PD2MiniSLBoosted

The D3PD2MiniSLBoosted code processes D3PDs into TRC ntuple format called MiniSL. TRC is a quite complicated infrastructure, so I'll describe just the most useful classes for a typical analysis user scope

- **SemileptonicBoostedSelection.cxx(.h)** code is where the event selection is implemented. Here it is possible to modify the cuts used or to decide at which selection step to save the file. In example if you use 10 cuts, you can decide to

save the events passing at least cut 5, maintaining information on subsequent cuts results. It is technically very easy, moving the line *“status=true”* along the list of different cuts.

- **MiniSLBoosted.cxx(.h)** and **MiniSLOrig.cxx(.h)** are the programs that fill the output ntuple branches. Here you have to intervene in order to add or remove info you need (or not).
- **AppBaseBoosted.cxx(.h)** is the code that manage the entire calling of other programs and the input parameters.
- **MiniSLRunBoosted.cxx(.h)** call programs that manage systematics and different electron or muon channels.

Some useful parameters available and already applied by the submission scripts follow:

- **boost** for boosted analysis.
- **noTrim** for no trimmed large R jets and to save all events at the first cut.
- **useClusters** to save cluster info .
- **sysOn** to produce systematics only.
- **useLoose** to produce also loose data.
- **useTruthParticles** to enable branches at truth particle level.
- **mcTruth_noCut** to save events just after the baseline cuts only(only truth variable saved).
- **allTrimFatJets** to save all trimmed fat jet and not only the one passing the corresponding cut.
- **mcType fullSim** if you are running on MC data.
- **dataStream Muons** if you are running on data muon stream.

- **dataStream Egamma** if you are running on data electron stream.

HistoMakerMasterBase

HistoMakerMasterBase is responsible for the histogram production, the event reconstruction, the template overlap selection and the particle level selection. This is a TRC integrated program structured in 3 main functions: `initialize()`, `execute()` and `finalize()`. Histograms are managed by the **SmartH** class in the namesake files `.cxx` and `.h`. It recognizes the histograms name by the convention used and described further on (so change here in case of name convention changes) and automatically set binning, title, axis and the other histograms properties. To add a new histogram follows this procedure:

- you need to declare a *SmartH* object in *initialize()*. It will be added automatically in a map ordered by the same name of the histogram;
- call the *Fill1D(...)* [or *Fill2D(...)*] class method in the *Fillall()* function passing the 4-vector of the object to be plotted (i.e. the hadronic top), the variable you want (i.e. the p_T) and the event scale factor.

The physical objects are managed by the **SmartVector** class in the namesake file. This is an overloading of the *root* `TLorentzVector` class with some more features; the most useful is the *IsFull()* class method to check if the 4-vector is ever been filled (useful for the global program variable management and memory saving). To get the original `TLorentzVector` object just call the *Vector()* class method.

In the *execute()* method the input ntuple (all charged in *initialize()*) are opened and the loop over the TTree calling *executeEvent()* is performed. Here you can add all the operation you need to do for every event, like selection, objects reconstruction and variable filling. In the *execute()* function, input file properties are recognized, the normalization factor and the event weight are evaluated. The Particle Selection is done here by the *Recluster()* method.

This program can also perform the **Top Template Overlap Method (TOM)** in order to tag top *fat jets*. To do that you need to set the proper parameter and the input template file list. This tagger can be used in parallel with the “standard” selection. The possible input parameters currently implemented are:

- **-local** if you run locally with a file with a different convention for file name and path so you need to define by parameter some setting otherwise automatic. To see how to define other local parameters take a look in LocalHistoMakerBo.sh. **This procedure is to be used just in emergency and test case, not for the full analysis!**
- **-f** input file list
- **-o** output histo file name
- **-onlyParticle** to save only *particle level* information and not parton level ones.
- **-particleCut** apply particle level cut. It's however done just for signal ttbar files.
- **-topAntitop** save top and antitop true information.
- **-noCut** for signal ttbar no cut ntuple processing.
- **-estimateFakes** for data driven fake estimation. To be used over loose samples.

- **-n [nEvent]** number of events to process.
- **-ovOn** activate the template overlap method and pass the template file list. **tested just for one template file**
- **-batchNum** number of batch job got from the number in the file list
- **-selectedEventLog** to create a list of runNumbers and eventNumbers of events passes particle or reco selection.

HistoMakerMasterBase can be inherited by simpler codes to accomplish specifics tasks, conserving all the functionalities already implemented in the mother class.

Unfold Master

The program performs the unfolding for a single kinematic quantity per time (i.e. hadronic top p_T) and produce the corresponding differential cross section distridution; different cross sections can be performed (relative, absolute, ...). Also different unfolding methods are available (MatrixInversion, SVD-Tikonov,). Kinematic quantities, unfolding method and cross section typology can be set via a parameters input file. Global parameters and variables are defined and initialized in the *globals.cxx(.h)* file(s). The main class is **UnfoldMaster** that manage all the process using a maps of **UnfoldBringer** objects, one for each input distribution (systematic, data, backgrounds, signal, ...). The cross section evaluation is done by the **CrossUnter** class. For now the systematic errors are summed in quadrature by the **SysCrafter** class, that makes you chose which systematic to take into account, evaluates absolute and percentage errors and plot them in a *tex* table. The unfolding is done by

the **UnfoldSmith** class by using the RooUnfold package. It is possible to choose the matrix inversion, the SVD, the bin-by-bin and the iterative Bayesian methods. The *MaicSkills.cxx(.h)* contains some general and useful standalone functions. All plots are saved in a common directory named `$PLOT_REALM/unfoldLast`.

As input files you will need the *standard* (or *nominal*) file, together with data, signal $t\bar{t}$ and backgrounds, and the *noCut* (or *generated*) file. You can also add the *loose* (*fake*) file for data-driven fakes background and all the generated systematic files. All files are produced by the HistoMakerMaster code. Many scripts are included in order to exec analysis programs with correct settings, to manage files in ordered directory, to merge histograms and to produce file lists.

Histogram naming convention

The histogram name should be composed by 5 or 6 specific tags separated by underscore characters. If you want to add a new option in any tag, you will need to modify the **SmartH** class and the unfolding code in order to let them recognize the new tag. The name structure follows:

tag1, , the event level of the object plotted in the histogram like reconstructed or truth object. The possible options up to now are:

- reco: for reconstructed quantities
- true3: for truth quantities at parton level
- particle: for truth quantities at particle level

- fake: for data-driven fake leptons quantities

In case of 2D histograms the same 1D histogram level will be reported for both the quantities plotted separated by the “VS” character like in *particleVSreco*.

tag2, , the object the histogram refers to like hadronic top or W boson. The most common options up to now are:

- hadTop: for the hadronic top
- lepTop: for the leptonic top
- diTop: for the ttbar system
- lep: for the lepton
- ecc....

tag3, [**optional**], to recognize some plot representing the same distribution but differing for something. For example the reco $t\bar{t}$ system could be reconstructed in our case using trimmed jet with d12 selection or the template overlap method; different tags must be used.

- boostOv: if template overlap method is used in reconstructed
- boostTrim: if trimmed jet d12 selection is used in reconstructed object
- leadingPt: if the truth particle level top is chosen as the leading p_T truth AntiKt10 jet
- leadingPt: if the truth particle level top is chosen as the truth AntiKt10 jet best matching with a truth hadronic top

tag4, , recognize the histogram type (1D or 2D). The only possible options up to now are:

- dist: for 1D histograms ;
- respo: for 2D histograms.

tag5, , is the quantities plotted in the histogram. The most common options up to now are:

- pt: transverse momentum
- m: mass
- eta: pseudo-rapidity
- ecc...

tag6, , the data type on the events, if it is a MC or real data event, form which kind of MC it came from. The only possible options up to now are:

- data: real data
- ttbarPow: signal ttbar PowHeg MC sample
- ttbar: signal ttbar McAtNlo MC sample diboson: diboson background MC sample
- stop: single top background MC sample
- qcd: QCD background MC sample or data driven one
- wjetsWhh: W+jets background MC sample
- wjetsWc: W+jets background MC sample
- wjetsll: W+jets background MC sample
- zjets: z boson + jets background MC sample
- zprime: z prime boson MC sample

This is an axample of a histogram name: reco_hadTop_boostTrim_dist_pt_ttbarPow

Installation

Some simple instructions are given for the whole framework installation, completely done by an auto-installer file.

- download from SVN the following package:

```
svn co svn+ssh://matteo@svn.cern.ch/repos/atlas-matteo/matteo/tags/BoostedRealm-01-00-00 BoostedRealm
```

- launch the installer

```
cd $HOME/BoostedRealm
```

```
source boostedRealm.inst
```

- *use the correct TRC tag;*

- Launch

```
./AdaptToGrid.sh ( to set the last version of the histogram maker programs )
```

```
./FastJetCompile.sh
```

```
./BuildTRC.sh (just to compile)
```

Every time you enter your system you need to setup the environment with the command.

realm_setup.

This allows to have a common structure on every system on which the analysis runs.

Bibliography

- [1] D. H. Perkins, Introduction to High Energy Physics.
- [2] S. Eidelman, et al. Physics Letters B 592 (2004)1-1109.
- [3]Perl et al., (1975).
- [4]Herb et al., (1977)
- [5]CDF Collaboration, F. Abe et al., Phys. Rev. Lett. 74, 2626 (1995),
arXiv:hep-ex/9503002 [hep-ex].
- [6]D0 Collaboration, S. Abachi et al., Phys. Rev. Lett. 74, 2632 (1995),
arXiv:hep-ex/9503003 [hep-ex].
- [7] Combination of CDF and D0 results on the mass of the top quark using
up to 5.8 fb⁻¹ of data, Technical Report FERMILAB-TM-2504-E, 2011.
- [8] CDF Collaboration, Conf. Note 9913.
- [9] D0 Collaboration, Phys. Rev. D 80 , 071102(R) (2009) [arXiv:0903.5525 [hep-ex]];
Phys. Rev. D 82 , 032002 (2010) [arXiv:0911.4286 [hep-ex]]; arXiv:1105.5384v1
[hep-ex].
- [10] ATLAS Collaboration, ATLAS-CONF-2011-100.
- [11] CMS Collaboration, JHEP 07 , 049 (2011) [arXiv:1105.5661 [hep-ex]].
- [12] T. Aaltonen et al., Phys. Rev. Lett. 103, 092002 (2009).
- [13] V. M. Abazov et al., Phys. Rev. D 84, 112001 (2011).
- [14] S. Frixione , S. Laenen , P. Motylinski , C. White , and B. Web-
ber , Journal of High Energy Physics 2008, 029 (2008).
- [15] C. White , S. Frixione , E. Laenen , and F. Maltoni , Journal of
High Energy Physics 2009, 074 (2009).
- [16] S. Eidelman, et al. Physics Letters B 592 (2004)1-1109.
- [17] L. Lyons, D. Gibaut and P. Clifford, Nucl. Instrum. Meth. A270, 110 (1988).
- [18] Precision Electroweak Measurements and Constraints on the Standard
Model, Technical Report CERN-PH-EP-2008-020, 2008.
- [19]ATLAS Collaboration, Observation of a new particle in the search for the Standard
Model Higgs boson with the ATLAS detector at the LHC, Physics Letters B, Volume
716, Issue 1, 17 September 2012, Pages 1-29, ISSN 0370-2693,
<http://dx.doi.org/10.1016/j.physletb.2012.08.020>.

- [20] CMS Collaboration, Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC, Physics Letters B, Volume 716, Issue 1, 17 September 2012, Pages 30-61, ISSN 0370-2693, <http://dx.doi.org/10.1016/j.physletb.2012.08.021>.
- [21] N. Kidonakis , Phys. Rev. D83, 091503 (2011).
- [22] N. Kidonakis , Phys. Rev. D82, 054018 (2010).
- [23] N. Kidonakis , Phys. Rev. D81, 054028 (2010).
- [24] ATLAS-CONF-2012-149.
- [25] CMS PAS TOP-12-006.
- [32] Luminosity Determination in pp Collisions at $\sqrt{s} = 7$ TeV using the ATLAS Detector in 2011, Technical Report ATLAS-CONF-2011-116, CERN, Geneva, 2011.
- [33] S. Agostinelli et al., Nuclear Instruments and Methods in Physics Research A 506, 250 (2003).
- [34] M. Mangano et al., JHEP 0307, 001 (2003).
- [35] S. Frixione , P. Nason , and B. Webber , JHEP 0308, 007 (2003).
- [36] J. Pumplin , D. Stump , J. Huston , H. Lai , P. M. Nadolsky , et al., JHEP 0207, 012 (2002).
- [37] G. Corcella et al., JHEP 0101, 010 (2001).
- [38] J. Butterworth and J. Forshaw , Journal of Physics G 19, 1657 (1993).
- [39] M. Aliev et al., Computer Physics Communications 182, 1034 (2011).
- [40] H.-L. Lai et al., Phys.Rev. D82, 074024 (2010).
- [41] S. Frixione , P. Nason , and C. Oleari , JHEP 0711, 070 (2007).
- [42] T. Sjostrand et al., Comput.Phys.Commun. 135, 238 (2001).
- [43] B. P. Kersevan and E. Richter-Was , Comput.Phys.Commun. 149, 142 (2003).
- [44] V. Abazov et al., Phys.Rev. D76, 092007 (2007).
- [45] C.-H. Kom and W. Stirling , Eur.Phys.J. C69, 67 (2010).
- [46] ATLAS Collaboration, ATLAS-CONF-2011-032,
<https://cds.cern.ch/record/1337782?ln=en>
- [47] M. Cacciari, G.P. Salam and G. Soyez, The Anti-k(t) jet clustering algorithm, JHEP 04 (2008) 63, 830 arxiv:0802.1189.
- [48] G. Zweig , CERN-TH-401 (1964).
- [49] S. Glashow , Nucl. Phys. 22, 579 (1961).
- [50] S. Hassani et al., Nuclear Instruments and Methods in Physics Research A 572, 77 (2007).
- [51] T. Lagouri et al., IEEE Transactions on Nuclear Science 51, 3030 (2004).
- [52] ATLAS Collaboration, ATLAS-CONF-2012-101,
<https://cds.cern.ch/record/1463915?ln=en>
- [53] <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/MUON/PublicPlots/2011/Dec/index.html>
- [54] <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/EGAMMA/PublicPlots/20120611/ElectronEfficiency2012/ATL-COM-PHYS-2011-783/index.html>,

- <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/EGAMMA/PublicPlots/20110512/CalibratedZee/ATL-COM-PHYS-2011-1637/index.html>
- [55] ATLAS Collaboration, ATLAS-CONF-2011-032,
<https://cds.cern.ch/record/1337782?ln=en>
- [56] <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/JetEtmissApproved2011PileupOffsetAndJVF>
- [57] J. Erdmann and O. Nackenhorst, Twiki page: KLFitter,
<https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/KLFitter>.
- [58] Keith Rehermann and Brock Tweedie. Efficient identification of boosted semileptonic top quarks 1327 at the LHC. *Journal of High Energy Physics*, 2011:1–27, 2011
- [59] M. Schmelling, *Nucl. Instr. and Meth. in Phys. Res. A* 340 (1994) 400–412.
- [60] G. D’Agostini, Improved iterative Bayesian unfolding, *arXiv:1010.0632v1*, 2010.
- [61] A. Hocker and V. Kartvelishvili, SVD approach to data unfolding, 1858 *Nucl. Instrum. Meth. A* 372 (1996) 469–481.
- [62] Jet Substructure at the Tevatron and LHC: New results, new tools, new benchmarks, A. Altheimer, S. Arora, L. Asquith, G. Brooijmans, J. Butterworth, M. Campanelli, B. Chapeau and A. E. Cholakian, et al., *J. Phys. G* 39, 063001 (2012) [*arXiv:1201.0008 [hep-ph]*]
- [63] J. Thaler and K. Van Tilburg, Identifying Boosted Objects with N-subjettiness, *JHEP* 1103 (2011) 015, *arXiv:1011.2268 [hep-ph]*.
- [64] J. Thaler and K. Van Tilburg, Maximizing Boosted Top Identification by Minimizing N-subjettiness, *JHEP* 1202 (2012) 093, *arXiv:1108.2701 [hep-ph]*.
- [65] J. M. Butterworth et al., Jet substructure as a new Higgs search channel at the LHC, *Phys. Rev. Lett.* 100 (2008) 242001, *arXiv:0802.2470 [hep-ph]*.
- [66] D. Krohn, J. Thaler, and L.-T. Wang, Jet trimming, *JHEP* 2010 (2010) 20, *arXiv:0912.1342 [hep-ph]*.
- [67] S. D. Ellis, C. K. Vermilion, and J. R. Walsh, Techniques for improved heavy particle searches with jet substructure, *Phys. Rev. D* 80 (2009) 051501, *arXiv:0903.5081 [hep-ph]*.
- [68] S. D. Ellis, C. K. Vermilion, and J. R. Walsh, Recombination Algorithms and Jet Substructure: Pruning as a Tool for Heavy Particle Searches, *Phys. Rev. D* 81 (2010) 094023, *arXiv:0912.0033 [hep-ph]*.
- [69] T. Plehn, M. Spannowsky, M. Takeuchi, and D. Zerwas, Stop Reconstruction with Tagged Tops, *JHEP* 1010 (2010) 078, *arXiv:1006.2833 [hep-ph]*.
- [70] L. G. Almeida, S. J. Lee, G. Perez, G. Sterman, and I. Sung, Template overlap method for massive jets, *Phys. Rev. D* 82 (Sep, 2010) 054034, [*arXiv:1006.2035*].
- [71] L. G. Almeida, O. Erdogan, J. Juknevich, S. J. Lee, G. Perez, and G. Sterman, Three-particle templates for boosted Higgs boson, *Phys. Rev. D* 85 (2012) 114046,

- [arXiv:1112.1957].
- [72] O. Gabizon, J. Juknevich, G. Perez, Y. Soreq, Measuring Boosted Tops in Semi-leptonic $t\bar{t}$ Events for the Standard Model and Beyond, arXiv:1311.2962v1[hep-ph] 12 Nov 2013
- [73] ATLAS Collaboration, Electron and photon reconstruction and identification in ATLAS: expected performance at high energy and results at 900 GeV, ATLAS-CONF-2010-005, 22 June 2010
- [74] ATLAS Collaboration, Identification of muon candidates in pp collisions at $\sqrt{s}=900$ GeV with the ATLAS Detector, ATLAS-CONF-2010-015, 26-Mar-2010.
- [75] ATLAS Collaboration, Jet energy resolution in proton-proton collisions at $\sqrt{s}=7$ TeV recorded in 2010 with the ATLAS detector, Eur. Phys. J. C, 73 3 (2013) 2306, 23 October 2012
- [76] The ATLAS collaboration, Measurement of top quark pair differential cross-sections in the $l+\text{jets}$ channel in pp collisions at $\sqrt{s}=7$ TeV using the ATLAS detector, ATLAS-CONF-2013-099, 15 September 2013
- [77] J.-F. Arguin, M. Franchini, O. Gabizon, A. Mengarelli, M. Negrini, A. Ovcharova M.D. Shapiro, F. Spano, W. Yao, Measurement of the top-antitop pair differential fiducial cross-section as a function of the top quark p_T in the boosted topology, ATLAS-COM-PHYS-2013-1093, November 20, 2013
- [78] P. Z. Skands, Tuning Monte Carlo generators: The Perugia tunes, Phys. Rev. D82 (2010) 074018, arXiv:1005.3457 [hep-ph].
- [79] ATLAS Collaboration, Jet energy measurement with the ATLAS detector in proton-proton collisions at $\sqrt{s}=7$ TeV, Eur. Phys. J. C73 (2013) 2304, arXiv:1112.6426 [hep-ex].
- [80] ATLAS Collaboration, Commissioning of the ATLAS high-performance b-tagging algorithms in the 7 TeV collision data. Geneva, Jul, 2011. ATLAS-CONF-2011-102.
- [81] ATLAS Collaboration, Calibrating the b-Tag Efficiency and Mistag Rate in 35 pb $^{-1}$ of Data with the ATLAS Detector. Geneva, Jun, 2011. ATLAS-CONF-2011-089.
- [82] ATLAS Collaboration, Improved luminosity determination in pp collisions at $\sqrt{s}=7$ TeV using the ATLAS detector at the LHC, Eur. Phys. J. C (2013) 73:2518, 10 July 2013
- [83] J. Campbell and R. Ellis, Nucl.Phys.Proc.Suppl. 205-206, 10(2010).
- [84] Valentin Ahrens, Andrea Ferroglia, Matthias Neubert, Ben D. Pecjak, and Li-Lin Yang. RG-improved single-particle inclusive cross sections and forward-backward asymmetry in $t\bar{t}$ production at hadron colliders. JHEP, 1109:070, 2011, 1103.0550
- [85] K. Agashe et al., Top quark working group report, arXiv:1311.2028v1 [hep-ph], 8 Nov 2013
- [86] V. Abazov et al (DØ Collaboration), Phys. Lett. B 693, 515 (2010), [arXiv.org:1001.1900].

- [87] T. Aaltonen et al. (CDF Collaboration), PRL 102 222003, [arxiv.org:0903.2850].
- [88] The CMS Collaboration, Measurement of differential top-quark-pair production cross sections in pp collisions at $\sqrt{s}=7$ TeV, Eur. Phys. J. C (2013) 73:2339, DOI 10.1140/epjc/s10052-013-2339-4
- [89] The ATLAS collaboration, Measurements of top quark pair relative differential cross-sections with ATLAS in pp collisions at $\sqrt{s}=7$ TeV
- [90] John M. Campbell and R.K. Ellis. MCFM for the Tevatron and the LHC. Nucl.Phys.Proc.Suppl., 205-206:10{15, 2010, 1007.3492}.
- [91] N. Cabibbo, Phys. Rev. Lett. 10, 531 (1963).
- [92] M. Kobayashi and T. Maskawa, Progress of Theoretical Physics 49, 652 (1973).
- [93] The LEP Collaborations, Phys. Rep. 427 (2006) 257,
<http://lepewwg.web.cern.ch/LEPEWWG>.
- [94] A.D. Martina, W.J. Stirlingb, R.S. Thornec and G. Wattc, Parton distributions for the LHC, arXiv:0901.0002v3 [hep-ph], 7 Jul 2009
- [95] The ATLAS collaboration, Measurements of top-quark pair differential cross-sections in the $t\bar{t}+jets$ channel in pp collisions at $\sqrt{s}=7$ TeV using the ATLAS detector TOPQ-2012-08, <https://cds.cern.ch/record/1470588/>, submitted to: Phys. Rev. D, TBD.
- [96] <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResults>